DOCUMENT RESUME

ED 141 377 TM 006 007

AUTHOR	Postlethwaite, T. N.; And Others
TITLE	Methodology for the Evaluation of Education Attainments; A Project of the IBRD and IIEP. IIEP Seminar Paper 2.
INSTITUTION	United Nations Educational, Scientific, and Cultural Organization, Paris (France). International Inst. for Educational Planning.
SPONS AGENCY	International Bank for Reconstruction and Development, Washington, D.C.
PUB DATE	75
NOTE	117p.; Paper presented at the International Institute for Educational Planning Seminar on the Evaluation of the Qualitative Aspects of Education (Paris, France, September 30-October 4, 1974); For related documents, see TM 006 003, 004, 005 and 006. Marginal legibility, pp. 34-35, 42-43, 45, and 47
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; Curriculum Development; Curriculum Evaluation; *Developing Nations; Educational Objectives; *Educational Planning; Educational Policy; *Educational Programs; Effective Teaching; Elementary Secondary Education; Enrollment; *Evaluation Criteria; *Evaluation Methods; Evaluation Needs; Formative Evaluation: Higher Education; *National Programs; *Program Evaluation; Summative Evaluation; Teacher Education; Teacher Evaluation
IDENTIFIERS	*International Bank for Reconstruction Development; International Institute for Educational Planning

ABSTRACT

The International Bank for Reconstruction and Development (IBRD), also known as the World Bank, has actively assisted in the development and improvement of educational facilities and systems in developing countries. A two-phase study was undertaken to improve the evaluation of World Bank projects. This report is an outgrowth of the first phase during which the major educational objectives of the IBRD-financed projects were studied and recommendations were made concerning appropriate methods and instruments for evaluating the attainment of the objectives of programs designed to improve student achievement and enrollment levels, teacher education, teacher effectiveness, and curriculum development. As illustrated in the report, the information obtained from program evaluation and experimentation will be used as integral elements in educational planning and policy development by the IBRD. (EVH)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED141377

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IIEP seminar paper: ②

METHODOLOGY FOR THE EVALUATION OF EDUCATION ATTAINMENTS, A PROJECT OF THE IBRD AND IIEP

T. N. Postlethwaite, B. H. Choppin,
N. L. Gage, A. Lewy, D. A. Pidgeon,
R. J. D. Chesswas

A contribution to the IIEP Seminar
on "The evaluation of the qualitative
aspects of education"
30 September - 4 October 1974

300 francs

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY
International Institute
for Educational Planning
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

TM006 007

INTERNATIONAL INSTITUTE
FOR EDUCATIONAL PLANNING
(established by Unesco)
7-9, rue Eugène-Delacroix, 75016 Paris

ⓒ Unesco 1975

The opinions expressed in these papers are those of the authors and do not necessarily represent the views of the Institute or of Unesco.

CONTENTS

Page

Foreword (iii)

Chapter 1. Background, purposes and initial procedures of the project 1

Chapter 2. The nature of the evaluation process 11

Chapter 3. The evaluation of enrolment 25

Chapter 4. The development and evaluation of curriculum 37

Chapter 5. The use of evaluation in World Bank projects: an illustration 68

Chapter 6. The evaluation of teaching effectiveness and teacher education 87

Chapter 7. Evaluation and experimentation as a component part of educational planning 95

FOREWORD

In June 1973, the International Bank for Reconstruction and Development agreed with the International Institute for Educational Planning that the latter Institute would conduct a research study to be entitled 'Project Evaluation Methodology - Education Attainment'. The research project would be conducted in two phases, and this report presents the findings of the first phase. The major objective of the first phase was a study of the educational objectives of the IBRD-financed education projects and a recommendation of the appropriate methods for evaluating achievement of these objectives. The study would focus on particular aspects of the educational projects: curriculum development, student achievement, teacher education, and teacher effectiveness. To this end, the following IIEP staff members and consultants co-operated in the work:

IIEP Staff Members

T. Neville Postlethwaite, Ph.D., ex-director of the International Association for the Evaluation of Educational Achievement (IEA), specialising in educational measurement, evaluation and statistical analysis and systematic curriculum development and evaluation.

Roger J.D. Chesswas, B.A., Junior Consultant to IIEP, Assistant to T. Neville Postlethwaite.

Consultants

Bruce H. Choppin, Ph.D., Principal Research Officer, National Foundation for Educational Research in England and Wales, specialising in all aspects of educational evaluation and processing and analysis of educational data.

Nathaniel L. Gage, Ph.D., Professor of Education and Psychology, Stanford University, specialising in research on teacher education and teaching effectiveness.

Arieh Lewy, Ph.D., Head of the Evaluation Unit of the Curriculum Center, Ministry of Education, Israel.

Douglas A. Pidgeon, Ph.D., Executive, Initial Teaching Alphabet Foundation, specialising in all aspects of the evaluation process.

It will be noted that, in the text, only quotations are referenced to their source. At the same time, readers familiar with evaluation literature will identify ideas from major writings in the field. A short bibliography of major reference works is given on pages 109/110 of this Report.

CHAPTER 1

BACKGROUND, PURPOSES AND INITIAL PROCEDURES OF THE PROJECT

In this chapter, we present briefly, on the one hand, the background and terms of reference of the present project, and basic principles underlying the present project on the other. The background arises from the nature of World Bank education projects to date. The basic principles are those that have been developed in the field of educational evaluation in general. Finally, this chapter presents some of the initial procedures of the project in obtaining a grasp on the nature of the objectives and the problems that would arise in evaluating achievement of them.

CURRENT BANK OPERATIONS: THEIR EVALUATION AND AUDIT

In the World Bank loans to date, the bulk of money has been spent on the building of new schools or the extension of existing schools. In the planning for such buildings and extensions, there have been detailed architectural plans, detailed plans for the equipment required, and detailed plans for the tendering procedures. Further to this type of objective, there have been a series of other objectives concerned with the expansion of enrolment, the growth of enrolment output, and the absorption of such output into the economy. Yet other objectives have been concerned with an increase in student achievement in the schools themselves, or through teacher-training colleges the improvement of teacher quality, and finally, objectives implying a contribution (by a World Bank project) to the economic, social, cultural and environmental development of the particular country.

The purposes of this study were set forth in the Terms of Reference dated 6 June 1973. These purposes are:

1. To recommend to the IBRD:
    - (a) Appropriate methodologies for the evaluation of important aspects of the Bank's education projects; and
    - (b) frames of reference for the formative and summative evaluation of future projects.
2. To develop practical methods and instruments for evaluating:
    - (a) Whether the intended student enrolments were achieved (considering, for example, distribution of students by programme, by sex, by ethnic groups, by catchment areas, etc.);

(b) whether the education objectives as stated in the appraisal reports, loan/credit agreements, side letters etc., were met; in this connection, it would be necessary to consider the project's impact in promoting reforms pertaining to education achievement; and

(c) whether the attainments of the students reached expectations (including measurement of cognitive, psychomotor, behavioural and attitudinal achievements, of drop-out rates, of repetition rates, of progression and examination rates, etc.).

3. To determine as implications for Bank education operations and practices, and possibly for Bank education policies, on:

(a) The extent to which the education objectives as defined in the past education projects permit appropriate quantified evaluation; and

(b) the extent to which education objectives are included in project agreements which are unsuitable for appropriate evaluation or are so generally formulated that assessment is impossible, in which cases such objectives might be amended or redefined in future project agreements to permit evaluation.

4. To recommend the appropriate machinery and instruments and the costs to organise and execute project evaluation, including a measurement of the fulfilment of educational objectives and of achievements in the project institutions. The intention should be to have such studies carried out as a routine practice during and after project execution. The research project should also estimate the additional costs for exploring the correlation between the attainments of the students and the internal (school) and the external (socio-economic background) factors which influence attainment.

The Bank has thus been preoccupied with different aspects of education project evaluation as is also indicated in the following clarifying statement from a letter of 14 August 1973 from Mats Hultin, senior adviser, Education Department of the IBRD, to Neville Postlethwaite:

"Education projects can be analyzed at four levels:

1.(a) Whether the intended construction, equipment, technical assistance was achieved (considering price, time, quality, use of local staff, materials, etc.);

(b) Whether the intended student enrolments were achieved (considering distribution of students by sex, by programmes, by catchment areas, etc.);

2.(a) Whether the education objectives as stated in the appraisal reports, loans/credit agreements, side letters, etc. were met;

(b) Whether the attainments of the students reached expectations (including measurement of cognitive, behavioural and attitudinal achievements, of drop-out rates, of repetition rates, of promotion between levels of examination rates, etc.);

3. Whether and how soon the graduates (school-leavers, drop-outs) found gainful employment in related fields; and

4. Whether the education projects contribute to economic, social, cultural and environmental development."

The relative emphasis of the current IIEP research project on these purposes was clarified in the same letter as follows:

> "The second level of analysis (analysis of education projects as to whether the education objectives were met and whether the attainments of students reached expectations) has the highest priority for the time being."

The project has proceeded accordingly, and the present report reflects the effort thus far to achieve these purposes.

The Bank's evaluation/audit of education projects at Level 1(a) has been particularly thorough. This can be seen from the full supervision reports as well as within-Bank audit reports. The criteria listed in 1(a) are typically used. Again, the evaluation/audit of education projects at Level 1(b) has been very thorough in some cases and can serve as a good model in its way. It is worth noting that the objectives for Levels 1(a) and 1(b) are typically spelled out in great detail, thus making the assessment of their attainment possible. On the other hand, the evaluation/audit at Levels 2(a) and 2(b) has had an ad hoc judgemental character to date. Evaluation at Level 3 by use of tracer studies has received a good deal of attention, particularly in latter years. Work at Level 4 has proved to be elusive, but efforts are currently underway, we understand, to improve the work at this level. Again, it should be pointed out that the objectives for Levels 2(a), 2(b) and 4 have been stated in vague terms, although for Levels 2(a) and 2(b), it is sometimes easier to infer what the more specific objectives might be.

Purpose of this report

The purpose of this report is to focus particularly on Levels 2(a) and 2(b) and to some extent on 1(b). The objectives of Levels 2(a) and 2(b) are of particular importance, since it is not enough simply to expand enrolments (for example, through increased capacity of new buildings). It is also necessary to increase the quality of what occurs within those buildings.

In view of the expressed policy that the World Bank will further emphasize the qualitative aspects of education in the future, it becomes increasingly important that the objectives implied by Levels 2(a) and 2(b) be carefully thought out and specified in such a way that evaluation/audit becomes more feasible. To identify the main qualitative (pedagogical) thrusts of the World Bank's interests in the loan agreements, the present project selected a sample of all appraisal reports to date and made an analysis of all pedagogical objectives stated in those reports. The objectives clustered into three main areas: teaching, curriculum and student achievement. Clearly, these three areas are interrelated. In many of the Bank loans, attempts were made to improve all three at the same time, but in others work was undertaken in only one or two of these areas. The bulk of the objectives concerned regular general, technical and agricultural secondary schooling and teacher-training institutions. Some of the objectives, however, were concerned with university education (including higher technical education), and primary and non-formal education (e.g., community education centres of the kind to be found in Tanzania). In some cases work was devoted to modernizing (from the country's point of view) the examination system; this kind of objective was taken up in this report in considerable detail, since it relates closely to aspects of curriculum development and student achievement.

The generalisability of proposed procedures

We realize that a gradual change is underway in the Bank's emphasis, leading toward a concern not only with secondary but also with other levels of education and with non-formal education. Although, because of the Bank's education lending so far, the bulk of the present report deals with regular in-school formal education, it should be noted that the procedures to be considered and followed would be much the same for other kinds of education, whether these be non-formal or university education. The procedures apply in general to developing as well as to developed countries. Thus, whatever the specific project or sub-project, the objectives must be carefully formulated. The types of behaviour considered to reflect the achievement of these objectives must be precisely specified. The necessary measuring instruments must be constructed with appropriate pre-testing, item analysis, etc. These procedures include determining the criteria or standards by which to evaluate the attainment of the objectives. The research designs and sampling procedures will have to be prepared and executed. Data will have to be collected, analyzed, interpreted and written up. It should perhaps be noted that much more expertise exists in the evaluation of activities in regular schools than in the tertiary or non-formal kinds of education. However, available technique would make it possible to undertake work of an illustrative kind in these other areas, given appropriate time and funds. However, the more unstructured the educational activity, the more difficult it is to identify and assess the relevant variables. Thus, it is clear that many variables influence behaviour in the cognitive, social-emotional, and psycho motor domains. Some progress has been made in recent years in identifying the more important variables, particularly the intervening variables, that should be taken into account in any evaluation exercise. This effort to identify significant factors in educational achievement will require much work in the coming years and particularly in developing countries.

Evaluation and experimentation as component parts of educational systems

The basic principle underlying this report is that any educational undertaking should have an evaluation component. This evaluation component should have three major tasks. The first task is to monitor the total system of education, as in determining differences in educational outcomes between students, schools, regions, etc. and in determining the independent variables (the input and process factors of the education system) associated with such differences. This monitoring makes possible the reallocation of resources that may be desired to reduce any inequality between groups in achievement or participation rates, etc.

The second task of the evaluation component is to assess different educational methodologies and innovations. This task would include the assessment of teacher skills, curricular materials, instructional procedures, and different methods of evaluation. We repeat that, in accordance with the priorities of the project as cited above, we shall not be concerned in this project with the evaluation of achievement of national goals as to economic, social, cultural, and environmental development. Nor shall we be concerned in this project with the appropriateness of any nation's goals.

This principle has led us to propose that evaluation and experimentation should be a component part of every educational system. Such a component will improve the efficiency of the system by providing continuous feedback on which corrective measures can be based. Such a component permits evaluating particular innovations (be they curricular or forms of teacher training, etc.) in terms of empirical evidence as to their 'workability' before they are implemented on a large scale. Finally, as a by-product, such a component will generate information concerning the systems which will assist the World Bank in determining the priorities for further financing of each system. In the last connection, we shall propose various modes of training that can increase rapidly the level of evaluation competence available in developing countries. Such increases would enable developing countries to take over the evaluation functions which must at present be initiated and financed by the World Bank.

Although any given World Bank project may touch upon only certain aspects or sections of a system, the ultimate object of World Bank activities is to help a particular country as a whole in its education, economic and social development. Therefore, the evaluation-experimentation component should eventually be extended to the system as a whole. Further, when evaluation is undertaken in a particular country at the inspiration of the World Bank, then the objectives of the evaluation are a matter of concern for both the country itself and the World Bank.

We would finally like to forward our conviction that the financial constraints that exist in the developing countries necessitate the establishment of appropriate mechanisms for the evaluation of both the quantitative and the qualitative aspects of the education system, which is generally the largest consumer of public funds. The current state of the art permits the establishment of such efficient evaluation mechanisms. The ultimate goal should be to incorporate such continuous evaluation mechanisms into every country's normal educational enterprise.

Initial procedures of the project

As a basis for the subsequent steps in the present project, an analysis of the objectives of recent World Bank education projects was needed. Such an analysis would provide what, as we indicate in detail below, must be the first step of any evaluation effort: the identification of objectives.

Accordingly, the first task undertaken in this project was to read, by the means of a microfiche reader/reproducer, the seventy-five sets of microfiches of the World Bank Black Books. Typically, each Black Book contains the following:

(1) A basic data sheet,

(2) The loan and/or credit agreement,

(3) The President's report,

(4) The appraisal report, and

(5) Side letters.

This reading of the sets of microfiches indicated that the objectives prior to about 1971 were mainly concerned with the expansion of enrolments in the existing system and with the relation between output and manpower demands. These stated objectives did not meet the requirements implied by the tasks of the project noted above, namely, to determine:

"- whether the education objectives as stated in the appraisal reports, loan/credit agreements, side letters, etc., were met.

- whether the attainments of the students reached expectations (including measurement of cognitive, psycho-motor, behavioural and attitudinal achievements)."

To obtain statements of objectives more concerned with educational achievement, teacher effectiveness, and the like, a sample of 36 of the more recent of the 75 sets of projects was chosen; i.e. this sample was biased towards the fiscal years 1971, 1972 and 1973. The abstracts (including not only the objectives but also descriptions and rationales of the projects) totalled about 450 pages. These were examined carefully, and the pertinent objectives were identified.

In mid-August 1973, eleven of these 36 abstracts were sent to the consultants to peruse, prior to a week's meeting to be held at the IIEP. In the meantime, the 289 identified objectives were extracted as direct quotations and classified in one document according to a set of categories of objectives which seemed to emerge. These categories and the percentages of objectives that fell into each of them are shown in Table 1-1.

The objectives in Categories 2-5 below, provided the central focus for the beginning of the research project. It will be noted in Table 1-1 that the objectives in these categories comprise about 61 per cent of all the 289 objectives considered. Thus this tabulation indicates that these 'qualitative' objectives - dealing with curriculum, teaching, achievement and matters relating to examinations, selection and guidance - constitute in these terms at least a highly important part of what World Bank education projects have been aimed at.

Table 1-1. The percentage distribution of enrolment objectives by separate category

Category of objectives	Number	%	
Enrolment objectives	54	18.7	
Curriculum objectives	73	25.3	60.9
Teaching objectives	35	12.1	
Achievement objectives	60	20.8	
Objectives relating to examinations, selection, and guidance	8	2.7	
Objectives relating to structure of the educational system	11	3.8	
Objectives relating to the implementation of reforms	19	6.6	
Objectives relating to special courses	10	3.4	
Miscellaneous	19	6.6	
Total	289	100.0	

From 3-7 September 1973, a group of consultants met at the IIEP, Paris, to examine the World Bank Projects in relation to the process of evaluation. Since there were severe inadequacies in the formulation of objectives (for evaluation purposes) the group prepared a brief report concerning evaluation in general and dealing in particular with objectives related to teaching, curriculum and achievement.

The specificity and clarity of objectives

Thus it becomes important to consider the degree to which these statements permit assessment of the degree to which the objectives have been achieved. And the conclusion must be reached that these statements are far from sufficiently clear or precise. They often amount to little more than rather vague declarations of intent. Objectives should state clearly exactly what standards are to be achieved and what degree of quality is required. In some project reports, it was possible to infer from a complete reading what may have been intended. In most instances, however, the necessary details were altogether unavailable. This lack makes the task of evaluation correspondingly vague. In a few cases where the appraisal report was traced back to the corresponding Project Identification Mission (PIM) report, no further information of the sort an evaluator would require was forthcoming.

It needs to be emphasized at the beginning, therefore, that the first major requirement of any evaluation is a clear and precise set of statements of objectives which indicate in detail what is to be achieved. With quantitative objectives this requirement is relatively easily met; there is little difficulty, for example, in specifying in detail the objective of building a number of secondary schools - specifying in such a way that a subsequent evaluation can determine whether the schools have been built in accordance with the required plans. The problem of producing precise and meaningful objectives in the area of enrolment is rather more complex, except where it is possible to state exact figures; as is shown in Chapter 3.

When objectives are concerned with qualitative aspects of education, however, the task becomes one of stating objectives in behavioural terms. This task is much more difficult and generally requires special training and practice. The difficulty arises from the need for clarity and observability. Statements such as "The new system ... will improve the curriculum" (Iran, 1970) or

"Curricula will be modernised" (Kenya, 1966), give no indication of what the improvements or modernisations actually mean. Usually the task of being more specific about the improvements intended requires stating what these mean in terms of the students' behaviours. It requires, for example, converting a generalised aim about the improvement of a curriculum to such a statement as "At the end of the course the students will be able to write a short essay demonstrating that they understand the structure and organisation of their system of government." Making such conversions requires training and skill. An effective evaluation must be preceded by a set of behavioural objectives prepared by someone expert in the task.

These impressions concerning the objectives of World Bank education projects and what was needed to make evaluation of these projects feasible guided the subsequent work of the project. After the Paris meeting in September, further analysis of needs and elaboration of plans were performed during the fall of 1973. The consultants met again in Paris in January 1974, to collate their independent work during the fall, to confer with Mats Hultin, who joined the group for that meeting, as to how the present report and plans could be improved for World Bank purposes, and to prepare the present final version of a report on the work and plans as of 31 January 1974.

CHAPTER 2

THE NATURE OF THE EVALUATION PROCESS

Because the evaluation process is centrally involved in the present project, this report provides here a brief characterisation of that process. This characterisation is intended to explicate the ways in which the evaluation of educational achievement has come to be carried out as a result of contemporary theoretical and practical work.

MAJOR STEPS IN THE EVALUATION PROCESS

In general, the process of evaluation consists of determining the degree and character of the value of something. In education, evaluation refers to the process of determining the degree to which the objectives of an educational activity, or enterprise, have been achieved. The activity may be a large one, in terms of its duration and scope, such as an entire course or even a curriculum for producing persons with certain kinds of professional or general skill. Or the activity may be relatively restricted, such as the work of a single instructor with his group of adults over a period of a week or an hour. In either case, the evaluation process determines the degree to which the 'objectives' - here used as a synonym of purposes, goals, aims, etc. - have been achieved.

As it has come to be practised in education, the process of evaluation has usually consisted of the following steps, here listed in the logical, and what is also typically the chronological order in which these steps are carried out. Different writers would vary in the specific ways in which these steps are clustered or grouped, but the essentials are well established.

Step 1. Identify the objectives of the educational activity or enterprise Often, perhaps, usually, the objectives are stated in terms of desired changes in the ways in which students can behave (their capabilities) or will typically behave (their habits or tendencies). This way of stating objectives has several advantages over such alternatives as stating objectives in terms of teacher activities or educational facilities. These advantages result from focusing on the ultimate objectives (student achievement of various capabilities and tendencies) rather than the various means (e.g., teacher activities or educational facilities) by which these ends will be sought. It is a distinction between ultimate and intermediate objectives. Once the intermediate objectives hav been attained and, for example, educational facilities have been constructed and teachers have been trained, evaluation must become concerned with whether their purposes - certain kinds of student achievement, capability, and habits - are being attained.

Thus, in the "appraisal of a fourth education project in Tanzania", it is stated in the summary and conclusions (paragraph v) that "Immediate needs of the education system are the expansion of teacher training, medical education, training of skilled craftsmen, and the integration of various types of rural education and training schemes (including) work-oriented education and training for youths and adults...". This kind of statement of objective should be converted, insofar as possible, into statements of desired changes in students' abilities and tendencies to behave in certain ways, e.g., perform such tasks as are entailed in various kinds of farming and skilled crafts.

Step 2. Specify the types of student behaviour that will be considered to reflect achievement of the objectives. In the cognitive domain, the student behaviour would consist of ability to remember, comprehend, apply, analyze, synthesize, or evaluate ideas, facts, concepts, principles, problems, and the like. These various kinds of abilities should be defined in observable terms. Thus, instead of resting with such terms as 'knowledge, comprehension, critical thinking, or grasping the significance of', one should attempt to use terms that imply observable behaviour, such as 'stating, recognizing, distinguishing true statements from false, matching, putting into one's own words, computing, naming, stating relationships between, or listing the consequences of'.

In the affective domain, the student behaviour would consist of various degrees of internalization of, and commitment to, attitudes and values considered desirable. Here again, an effort to maximize observability should be made. Thus, instead of such terms as 'appreciate, have an interest in, respect, etc.', one might use such terms as 'on his own initiative, he seeks out, tries, approaches, speaks favourably of, spends his own money for, pays attention to, etc.'.

In the psychomotor domain, the student behaviour readily takes such observable forms as ways of moving large and small limbs, communicating and expressing non-verbally, and motor aspects of ways of speaking and writing.

Step 3. Construct situations in which the student will be required or expected to demonstrate the desired ways of behaving. The situations may range from simple and short questions (or 'items') that will elicit a kind of knowledge to more elaborate problems that will call for a higher mental process (e.g., synthesizing or evaluating). The situations may be verbal, numerical, mechanical, spatial, esthetic, social, or whatever - the main criterion being relevance to the objectives. The situations may be contrived entirely or the kind that occur regularly in 'real life'.

In devising such situations, it is often desirable to make surveys of representative samples of the real-life conditions and circumstances under which the desired ways of behaving would be exhibited. Thus, when it comes to certain kinds of farming skills, the objectives should be defined on the basis of surveys of the kinds of soils, crops, terrain, and the like, in which the farming will be done. For various kinds of skilled crafts, surveys should be made of representative samples of the kinds of carpentry, wood, joints, structures, etc., that a carpenter would need to deal with in the work for which he was being trained. Similarly, the kinds of habits of accuracy, neatness, punctuality, and dependability entailed in the successful performance of the work would be specified. All of the specifications should be based, insofar as possible, on carefully designed surveys of representative samples of the real-life situations in which the student or trainee will subsequently work.

For certain kinds of work, various kinds of general knowledge, intellectual capabilities, and habits are required. The definition of observable behaviours that will cover these kinds of general requirements of living and working in a society should be based on a broad knowledge of the social and cultural characteristics of that society. Here we refer to such matters as literacy and numeracy, health habits of various kinds, and other kinds of behaviour necessary to fitting in and contributing to the society.

Step 4. Determine the criteria or standards that will be used to assess the value - correctness or desirability - of the behaviour elicited. In simple paper-and-pencil tests, this step consists of determining the 'scoring key', i.e., the list of correct answers. In more complex or non-symbolic situations, such as those in which a real-life performance is to be evaluated - e.g., a dance, a lathe operation, a meal preparation, or a discussion-group participation - this step consists of (a) a list of important dimensions of the performance and (b) an accompanying set of scales for use by observers, measurers, or judges of the performance.

Some samples of non-verbal tests are shown in Figures 2-1 and 2-2.

The student is shown a setup consisting of a hack saw, a vice holding a piece of work, and a saw cut started. He is asked, "Which one of the following is a correct statement?"

(a) The hack saw blade is in the frame backwards.

(b) A finer toothed blade should be used on this job.

(c) The piece protrudes too far from the vice.

(d) One should saw on the other side of the vice.

(Note that any four of the responses could be made the desired response.)

Which one of the following statements explains why this lathe tool bit cut rough?

(a) The feed is too high.

(b) The speed is too fast.

(c) The work is too loose between centres.

(d) The cutting edge is too high.

Figure 2-1. A test situation for a shop course

Food Score Card

Plantain

Appearance			Score
1	2	3	2
Shrivelled		Slightly Moist	

Colour			Score
1	2	3	3
Pale or Burned		Well-Browned	

Moisture Content			Score
1	2	3	1
Dry		Juicy	

Taste			Score
1	2	3	1
Raw, Tasteless or Burned		Flavour Developed	

Figure 2-2. A rating scale for evaluating a food product in a home economics course. The dimensions of cooked plantain to be evaluated are its appearance, colour, moisture content and taste. For each of these a three-point rating scale is provided.

Needless to say, the criteria to be used should be determined by the values, customs, and standards prevalent in the culture of the society in whose education system the evaluation is being performed. Tests, criteria, and standards appropriate in one culture will in many instances be inappropriate in another. Only curriculum experts in any particular country can judge and ensure this appropriateness of criteria and standards.

Step 5. Apply the measuring instruments - the tests, rating scales, situational performance tests, etc. - to the students whose achievement of the objectives is being determined. This step refers to the administration of the test. It can take the form of individual or group testing under highly or loosely standardized conditions, depending on what is necessary and feasible for the purposes of the evaluation.

Individual testing is, of course, much more expensive than group testing. It requires much more time on the part of the examiners, and frequently it is more important in individual testing that the examiner be highly trained. Group tests, by definition, can be given to groups ranging in size from 2 to 2,000 students at a time depending on the size of the auditorium and the number of proctors available. The cost of testing each student is thereby reduced materially.

Most individual tests usually require hand scoring by experts, whose time is expensive. Group tests, on the other hand, can be scored by clerks or by machines.

Thus, in general, group testing is to be preferred except under special conditions. These conditions are those in which the testing must be carefully adjusted to the child or adult being tested. If highly idiosyncratic behaviours are being looked for - behaviours of the kind that might be exhibited by exceptional children or adults (either retarded or highly gifted), or are to be examined for signs of creativity and originality - then individual testing may be necessary. But, for the vast majority of educational objectives for the vast majority of students, group tests will serve. Individual tests can take the form of oral examinations and interviews, and their high expense is justifiable when the importance of the examinee or the position to be filled is correspondingly high. But for the purposes of selecting a considerable number of persons from a large number of applicants, the less expensive group testing procedures usually turned out to be at least equally valid and much less costly.

Step 6. Score the behaviour - the responses or performances - elicited by the measuring instruments. The score will take such forms as 'number of correct responses' or 'total rating on the various dimensions' or 'accuracy of the product prepared in relation to specified attributes'. The score should tell the degree to which the objectives embodied in the questions or items of the measuring instrument have been achieved by the student. The degree to which the examination or test does yield accurate information on achievement depends on its reliability and validity.

Reliability is the degree to which the measuring instrument yields consistent results, regardless of what it is actually measuring. Thus, if a test yields the same score from one occasion to another - whether the occasions are separted by minutes, hours, days, or weeks - it is said to be reliable. But if memory is likely to make the student respond the same way to the same test on two different occasions, the reliability of the test is better estimated by using two equivalent forms of the same test. Then the agreement between scores on the two equivalent forms - an agreement which is often measured by means of a coefficient of correlation - is used as an estimate of the test's reliability. Or, reliability can be estimated by determining the degree to which two equivalent halves of the same test yield results that agree, or correlate, with one another. Finally, reliability can be estimated in terms of 'internal consistency', or the degree to which the items of a test yield scores that agree, or correlate, with one another. Statistical formulas have been developed and are well established for use in estimating reliability in these various senses. They are used in estimating an important characteristic of the measuring instrument - the degree to which it yields measures that are 'accurate' apart from the degree to which they measure what the test is actually intended to measure.

Validity refers to the latter characteristic - the degree to which the test measures what it is intended to measure. What it is intended to measure is, in the case of achievement tests, defined by the educational or instructional objectives whose achievement the test is designed to measure. For most achievement tests, validity can be determined primarily by inspection of the test's content and of the psychological or behavioural processes that it elicits from the student. Thus, if a test is intended to measure higher mental processes, such as reasoning ability or the ability to synthesize information of various kinds, but inspection reveals that it calls upon the student merely to recall or recognize information that he has received from books or lectures, then the test would be judged to be low in validity. Another example of low validity would be that of the test which is designed to determine the student's attitudes toward, or appreciation of, some of the literature he has been studying. If the student knows that the teacher or school desires him to express favourable attitudes, and if he knows that his responses will not be anonymous and that unfavourable responses will redound to his disadvantage, then again the test can be judged to be low in validity. It is simply not measuring what it was intended to measure.

The validity of tests used for selection or research purposes can take different forms from the judgemental procedures appropriate for achievement tests. When a test is used for selection purposes, its predictive validity is estimated by the degree to which scores on the test correlate with subsequent measures of criterion performance on a job. Thus, tests used to select draftsmen would have high predictive validity if scores on the test correlated substantially with ratings of the draftsmen obtained some months later from their supervisors. In short, the tests would have succeeded in predicting success on the job and would thus be considered to have predictive validity. If the criterion measures of success on the job are obtained at the same time as the test scores, the resulting correlation between the two sets of measures is termed the test's concurrent validity. Thus, a test used to measure student attitude toward a given author which correlated substantially with whether or not the students borrowed books by that author from the school library would be considered to have high concurrent validity.

Step 7. Judge the degree to which the score obtained by the students reflects achievement of the objectives of the educational activity or enterprise. Two major approaches to making such judgements are termed the norm-referenced and the criterion-referenced approaches. In the former, each student's score is compared with that of other students constituting a norm group. Such a norm group might be 'other students who have just taken the same course', or it might be 'a representative sample of students of the same grade-level throughout the nation'. The given student is then found to have a certain percentile rank in relation to the norm group, i.e., to equal or exceed a certain percentage of the students in the norm group.

The success of a norm-referenced test is judged by the degree to which it discriminates among students. The purpose of the test is to spread students out - to put them in a rank order ranging from the highest achieving to the lowest. A test that does not discriminate in this way is judged to be unsuccessful in creating variance, and hence it cannot be reliable or valid, if norm-referenced approaches are being used. Norm-referenced measurement imposes a kind of competition among students. Each student's achievement is evaluated primarily by being compared with that of the other students in his class, school, or community. By this approach, some students are forced to be inferior, because some students must by definition be below average; indeed, half of the students must always fall below the median, of course, and suffer the corresponding implications of inferiority.

For many years, it was considered necessary to use the norm-referenced approach because it was assumed that no other way of judging educational achievement - other than by comparing that of one student with that of other students - was possible. More recently, it has been realized that achievement measurement can be referred to objectives rather than to other students. If it is possible to set up educational or instructional objectives, and to define observable behaviours that indicate achievement of those objectives, then it is possible to evaluate achievement by eliciting and evaluating those behaviours directly, without reference to the performance of other students.

In criterion-referenced interpretation of test scores, the judgement is made directly with reference to the pre-specified types of behavioural or performance objectives of the educational activity. If the objectives referred to 'ability to read a typical newspaper article', then the interpretation of the student's score refers directly to whether such an ability has been demonstrated. If the objective referred to 'ability to produce a steel cylinder', with dimensions of a certain accuracy, on a lathe, then the criterion-referenced interpretation refers directly to whether such an ability has been demonstrated.

Other examples could be found in the objectives of programmes for training vehicle drivers, or subsistence farmers, etc. So-called 'Minimal learning packages' intended to develop basic skills can be evaluated by criterion-referenced approaches.

Norm-referenced evaluation has been used frequently in the past. It is closely linked to 'grading on the curve', which requires that only certain percentages of students be regarded as excellent, good, fair, etc. Criterion-referenced evaluation is coming to be used more and more as newer approaches to education are adopted - approaches such as programmed instruction, individualized instruction, self-paced instruction, and mastery-learning approaches.

In criterion-referenced measurement, the shape of the ideal or desired distribution of test scores is not the traditional 'bell-shaped' or any other distribution that shows a great amount of variance among students. Rather, the ideal distribution is one that shows the maximum possible number of students achieving the maximum or almost-maximum number of instructional objectives. That is, in criterion-referenced measurement one hopes that scores will pile up at the high end of the distribution. To the degree that such a goal is not achieved, one looks toward improving methods of instruction rather than resigning oneself to accepting inevitable individual (and perhaps hereditary) differences in aptitude or intelligence among students.

The basic distinction between norm-referenced and criterion-referenced measurement has implications for various other aspects of achievement evaluation. As already indicated, in criterion-referenced testing the variability yielded by the test is irrelevant, while in norm-referenced testing it is desirable and even crucial. Similarly, in criterion-referenced testing the main criterion is relevance to, or significance for educational objectives, or the criterion of successful educational achievement, regardless of either the 'difficulty' or the degree of 'discrimination' achieved by the test question or item. Similarly, the reliability of a test, insofar as it depends on the test's producing variability between students, may become irrelevant. If all of the students get perfect or nearly perfect scores, the test will have low reliability in the usual senses that depend on the variance among test scores, and yet be highly reliable and valid in the sense that the student's achievement does or does not consistenly reflect achievement of the objectives.

The formative and summative functions of evaluation

It is regarded as axiomatic that one of the major purposes of any World Bank project is to improve the quality of education in the country in which the project is operating. Evaluation should serve this overall purpose - the improvement of quality of education. Contrary to what is sometimes thought, evaluation is not an end in itself; its main purpose is not merely to provide information on whether or not a particular project has achieved the kind of success intended. Rather, in the long run, evaluation should help those responsible for the development and execution of the project to ensure that success is actually achieved. There are, however, numerous occasions when circumstances will only permit the assessment of final success, and such evaluation can also often produce information of value.

Misunderstandings about the function of evaluation arise from the failure to distinguish between summative and formative evaluation. The former comes at the end of an educational activity or project, while the latter occurs continually during the activity or project, with the aim of producing information which can be fed back to ensure that the appropriate aims and objectives are being attained. The information fed back from formative evaluation may necessitate changes in any of the first four steps listed earlier. They may call for changes in the situation or procedures developed in order that specific objectives will be achieved, or in the determination of the criteria or standards used to assess the value of the required behaviour. There are occasions, however, when modifications may be required in the objectives themselves.

In theory, whether an evaluation is to be summative or formative, the objectives to be achieved should be clearly stated at the outset of a new project. In practice, it is often the case that some summative evaluations are added on to a project almost as an afterthought. In such instances, formulations of objectives may not be produced until the project is almost completed. The danger here is that the evaluation may then become more concerned with those objectives which have in fact been achieved, rather than with those which the project was originally intended to achieve.

The uses of summative evaluation

Two kinds of circumstances can be identified in which summative evaluation may legitimately be carried out. The first is in a project, usually a short term one, in which the circumstances will hardly permit of any changes being introduced during the course of the project. An example would be a project which required the building of a number of additional secondary schools, or further teacher-training colleges, without any initial specifications that the curricula in the new establishments should be changed to meet new requirements. The summative evaluation would then be largely concerned with quantitative aspects, or perhaps with ensuring that there were no differences in the average levels of achievement of students in the new institutions as compared with those already existing.

In making a summative evaluation, it is often desirable to have a baseline against which the results of a given educational activity can be measured. Such a baseline can be obtained from the performance of a 'control group' - a group of students who have not received the kind of educational experience or training that is being evaluated. Sometimes such a control group received no training at all. In other instances, the control group may receive the traditional or regular kind of training that one is trying to improve upon. Ideally, the experimental group - the one that receives the new kind of training - and the control group - the one that receives the old kind of training or none at all - are 'randomly equivalent'. That is, students in these two groups are assigned at random from the total group, by some mechanism such as a table of random numbers or the tossing of a coin. Such randomization, if it is feasible, and if the control and experimental groups are large enough, is sufficient to ensure that all other possible explanations of any significant post-training differences between the two groups cannot be attributed to extraneous factors, such as differences in age, social class, home background, or whatever. For randomization makes the two groups non-significantly different in all conceivable factors other than the experimental variable - the difference between the new and old kinds of education, curriculum, or instruction.

Sometimes it is possible to introduce additional refinements by using statistical methods to adjust for whatever differences between the experimental and control groups may remain even after random assignment, or randomization. Thus, if the two groups are found to differ in scholastic aptitude, even after randomization, this difference can be used to adjust the post-instructional achievement test scores so as to eliminate such aptitude differences as an explanation of the achievement differences. If the adjusted achievement scores still differ significantly, i.e., to a degree greater than can be accounted for by chance fluctuations in random sampling, then it must be inferred that the instructional differences made a genuine difference in achievement that cannot be attributed to differences in aptitude.

The second circumstance in which summative evaluation may be legitimately required occurs at the end of a project during the course of which formative evaluation had been carried out. The primary aim of formative evaluation is to ensure, through changes brought about by the feedback process, that the originally stated objectives of a project are being achieved. In many instances, however, the feedback process causes changes to be made in the initially specified objectives, because some may prove to be unattainable in their original form. Where this happens, a summative evaluation should report on the extent to which both original and modified objectives have been achieved. An instance of this second type of summative evaluation might occur in a project which required "revisions of the secondary school structure and curriculum" in order to improve "the standard of secondary school education" (Ethiopia, 1966).

Any originally planned objectives for bringing about the improvements required might well have undergone a revision as a result of a feedback from a formative evaluation carried out during the course of the project.

It has already been emphasized that the translation of general goals into appropriate sets of behavioural objectives forms the first important step in the evaluation process. Detailed behavioural objectives take time to be developed, however, even by an expert. It is not reasonable to expect that in the time available they will readily flow from the pen of a member of a Project Identification Mission, for example, whatever expertise in this area he may possess. It is suggested that at this stage and in the appraisal reports much more detail about intended qualitative objectives be included. Then a set of detailed behavioural objectives can be produced by the time a project is ready to get under way.

The uses of formative evaluation

Since the major aim of formative evaluation is to improve education, the collection of data on the achievement of any objective should be undertaken as soon as appropriate after the commencement of a project. The results of this assessment should be fed back immediately to those concerned with the project's development. If these results show that the stated objectives are indeed being achieved, the project can continue along the lines already adopted. The results may reveal, however, that the improvement sought, or the achievement of the standards desired, is falling below expectation. Then steps can and should be taken to change the project development plan in a direction that subsequent evaluation will reveal to be more appropriate for the achievement of the stated objectives. Only a continuous evaluation procedure of this kind can ensure that the objectives will be achieved and the improvement will be brought about. Such formative evaluation does, in fact, occur in the construction of hardware (e.g., school buildings) when Bank architects inspect and modify construction projects in process. It is here being recommended that it also occur in the conduct of educational programmes and courses.

It will be seen from the above that the usefulness of summative evaluation is somewhat limited. It is rarely possible, with such evaluation, to make judgements or conclusions about any improvement that may have taken place since the inception of a project. Further, there is no opportunity to effect a change if the summative evaluation demonstrates that particular objectives have not been achieved.

This brief introduction to the process of evaluation implies a number of important questions about the evaluation of projects already under way, but not yet completed. It also implies the need to consider the whole question of evaluation at the very outset of any new projects. Later sections of this report deal with both these implications.

CHAPTER 3

THE EVALUATION OF ENROLMENT

This chapter deals with the enrolment objectives of World Bank educational projects, the concepts of stock and flow statistics, and the ways of measuring the achievement of enrolment objectives in terms of these concepts.

The enrolment objectives in the World Bank Black Books

One principal component of all the credit and loan agreements that have been examined in detail is the 'construction' of an educational establishment - a general secondary school, a teacher-training college, an agricultural college, an educational research institute or whatever. Four types of construction activities have been noted:

(a) building schools of exactly the same type as existing schools;
(b) expansion of some (or all) of the existing schools;
(c) creation of new schools similar but not identical to existing schools; or
(d) creation of a completely new type of school (with which there is no experience in the country).

This kind of construction objective logically leads to two kinds of enrolment objective to be found in the current and past Black Books - those concerned with actual school-related enrolments and those concerned with enrolments related to factors external to the school system itself. Within these two broad categories, it is possible to identify sub-categories of enrolment objectives.

Enrolment objectives related to factors internal to the school

(a) Simple increase in enrolment;
(b) the increase in enrolment related to the existing stock of students or projected stock of students;
(c) the reduction of high numbers of repeaters and drop-outs;
(d) the increase in enrolments and increase in output;
(e) the distribution of enrolments by sex.

Enrolment objectives related to factors external to the school

(f) Enrolments and the school-age population;
(g) regional distribution of enrolments.

The above classifications of types of schools and enrolment objectives are, in comparison with other objectives to be found in the Black Books, written in relatively precise and specific terms. One finds real, sometimes proportionate, figures readily amenable to evaluation. One must remember that, since an Appraisal Report cannot be longer than twenty pages, more specific information cannot be found in the text, and hence is placed in an annex. Furthermore, working papers accompanying each Appraisal Report give background information beyond that found in the Black Books. Also available are supervision reports, but since the architectural supervision is by far the greatest component of such a report, there is little else available in the way of commentary on, or evaluation of, non-architectural aspects such as enrolment, curriculum, teacher training, and so on.

Despite all this material, a very small amount of which has been examined in detail, the information on enrolments is really no further augmented, except for the occasional flow chart or distribution of enrolments by level and type of education, by year, and sometimes by sex. Thus, as they stand, the objectives on enrolment, which are for the greatest part concerned with a simple increase in capacity or addition in the number of school places (and readily amenable to evaluation by the simple counting of heads in the year 'y' of completion) do not contain the specifications that the World Bank asks for in its Terms of Reference for the present project; to quote from the Terms of Reference: "The study should concentrate in developing practical methods and instruments for evaluation (a) whether the intended student enrolments were achieved (considering, for example, distribution of students by programs, by sex, by ethnic group, by catchment areas, etc....)". This statement has been interpreted as indicating that the methods for evaluation should be concerned with measuring as much of the stock and flow statistics as are available. For not only do the Bank objectives examined deal with an increase in the stock of students in school, but they also mention such factors as output, promotion, repeater and drop-out rates, enrolment ratios, and so on.

One of the most important factors in the specification of objectives is that of differentiation between various contextual situations. For within the educational system itself, the specifications could well relate to factors at the national level, regional level, or even the catchment area. There is also the distinction between projects already begun and those that are just beginning. And probably most important of all is the distinction between Bank schools and non-Bank schools. All such 'levels' of interest must be spelled out in detail in order that appropriate measures can be identified and applied.

The concepts of stock and flow and the links between the two concepts

The stock of students in the educational system at a specific time is the result of a complex series of flows of students over time. Children enter schools at varying ages; some move from one grade to another; some repeat; some enter another level; some move from one type of school to another, and even some drop out (some re-enter) along the many paths possible. From the numerous possible flows of stock statistics, the actual statistical needs (as opposed to availability) are complex and must be flexible and complete to meet different computation demands.

The fundamental requirements are:

(1) the existing stock;

(2) the flow of students into, within, and out of the system.

It is fairly safe to state that stock statistics are typically easier to find than the types of data needed for the analysis of flow statistics, although crude computations can be made from a by-year series of stock statistics.

The basic possible statistical information can vary along a continuum from specific data about individuals to data at the aggregated national (or world-wide) level, according to the uses of, and computations to be made from, the available data. With few exceptions, the statistics in developing countries are never at the individual level and are often the aggregated national whole. In this respect, any immediate evaluation of the quantitative objectives outlined in the Bank Appraisal Reports will, in the first analysis, be restricted to the use of more aggregated statistics.

In the past, Bank investments have been concerned with expansion of enrolments more in conjunction with national development goals, and hence global national statistics were sufficient (and typically all that were available). Recently, however, the interests and the concerns of the World Bank have become more focussed on the rural areas of developing countries, and the disparities between different regions in any one country - not only urban and rural areas but geographical and ethnic regions as well. In this respect, therefore, data must now be collected in order that such disparities can be assessed, and that varying situations in rural areas can be examined at whatever level of aggregation of disaggregation of data is so required. Therefore, information concerning school statistics, and certain other data or factors external to the school, must be acquired at the national level, at the regional level, and by socio-economic group. Where sample surveys are being conducted, information on each individual must be collected. What have been perhaps the most feasible data to obtain, excluding a survey are:

(1) Student statistics: by level, type, age, sex, grade, region and socio-economic group. For flow analysis: entrants/intake leavers and output (both terminal and non-terminal), repeaters and promoters.

(2) Teacher statistics: according to the same bases as those used for student statistics.

(3) Statistics on school: by level and type, by region, the number of grades offered, the number of classes.

Furthermore, in the determination of student statistics, as has been mentioned, the socio-economic origin of the student would be a necessary datum, for in the analysis of school enrolment in relation to catchment area, such information is necessary. So, the total population must be broken down from the national level to the smallest possible administrative unit (e.g., district, village, etc.). It must, however, be realized that, at such a micro level of analysis, apart from data collection problems, 'geographical' units can, with further sophisticated study, be broken up into social strata, and so on. Such sociological surveys, however, require explanation and meaning well beyond the scope of purely statistical regional or ethnic disparities.

MEASURES FOR THE EVALUATION OF ENROLMENT OBJECTIVES

In this section, the various desirable computations will be listed and their intended measures explained. It will be assumed that the actual methodology for the computations is understood. Therefore, the principal task here will be to identify all the possible kinds of measures needed.

In the preceding section, reference was made to formative and summative evaluation. The same types of basic principles apply when evaluating enrolments. There is, however, a difference because sometimes base data on enrolments are available for projects that have been completed, whereas for student achievement the data on completed projects are not available. In this respect, therefore, some quasi-formative evaluation of enrolments can be undertaken, where year-by-year information on enrolments has been collected during the project construction or expansion. Although nothing can be done, as is possible within say curriculum development, about objectives that were not met during the project implementation, critical examination of the enrolment data is possible, and further 'mis-direction' can be checked. Also where baseline data exist, comparisons with the base year can be made to obtain an idea of the impact and relative importance of the project. This possibility will become more apparent as measures for evaluating enrolments are identified and discussed.

One final issue that must be dealt with before considering these measures is that of what base data one can realistically assume to be avilable for both past, present, and future projects. The situation is a dynamic one, of course, and sophistication in the collection of statistical data on schools is increasing in developing countries; so the future is promising. Past data are not, however, so complete nor sophisticated. Hence summative types of evaluation of past projects will be crude and not always altogether accurate. Typically, base data of former years are national school statistics, perhaps broken down by level, but age, sex, and regional statistics are available less and less often the further in time one goes back.

Stock analysis

Typically stock analysis is found only occasionally in an Appraisal Report, and sometimes in the accompanying background documents. The stock analysis includes analyses of the different proportions of enrolments by level and type of education, age, sex, and grade. With the suggested further examination of such enrolments by region and by socio-economic origin, the stock analysis will give an interesting static picture of the situation being examined.

There are further stock analyses which may be undertaken by continuing different factors which can give other interesting means of evaluation enrolments, like for example the following graphic distribution of enrolments by age and grade combined which shows the age variation in any one grade over a whole cycle.

Another example is that of the analysis of the distribution of school buildings and what they offer. The availability of facilities is an important factor in determining the opportunities of children to go to school. School mapping is a technique which has refined the statistical analysis of school buildings, and the distribution of the supply of educational facilities. Since the Bank's investments have been concerned primarily with the construction and extension of schools, the comparison of the different types of school will give an idea of the relative importance of Bank activities and the way in which they are important.

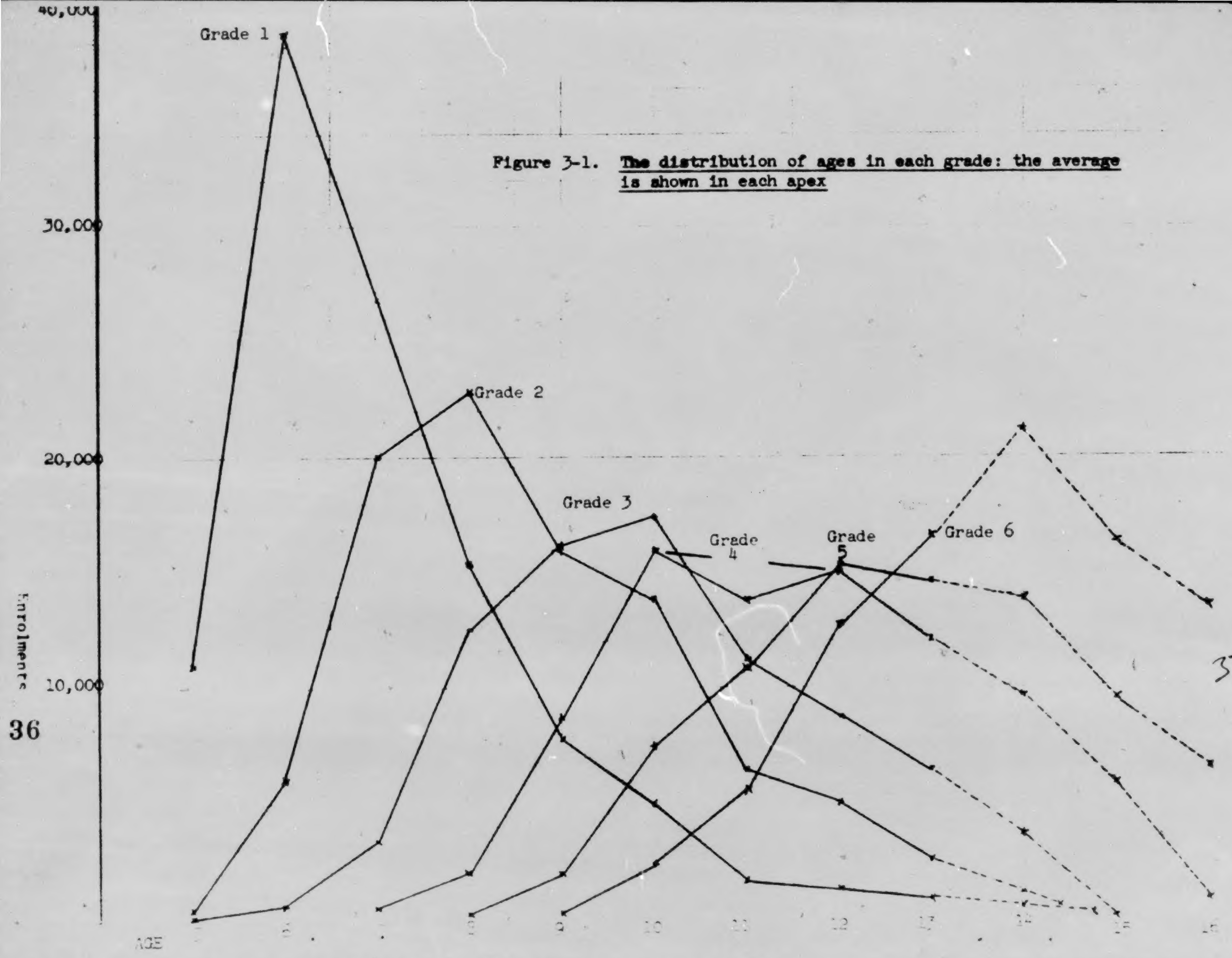


Figure 3-1. The distribution of ages in each grade: the average is shown in each apex

The characteristics to be examined are the following:

(a) Geographical distribution of Bank and non-Bank schools by level and type of school.

(b) A comparison between Bank projects and existing schools as to:

  (i) The number of grades offered by level and type of schooling.

  (ii) The number of classes offered by level and type of schooling.

  (iii) The size of classes offered by level and type of schooling.

Such data would show the degree of completeness of each school, and as experience shows in rural areas of developing countries, especially in Latin America, there are many incomplete schools: those which offer only three or four grades; and even those which offer few incomplete grades housed in one classroom. Such a phenomenon is important to remember in flow analyses also since this will force children to leave the cycle, provided they cannot find places elsewhere, and to be counted as a drop-out.

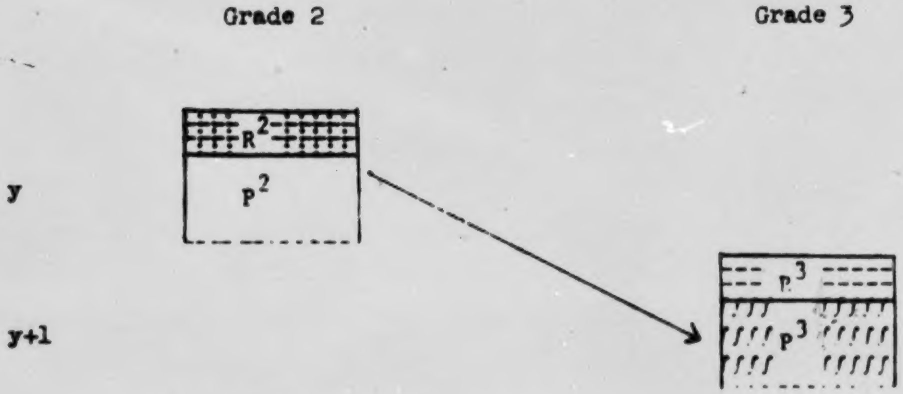
Such combination of indicators gives the educational system a broader description and allows deeper analysis of the data provided.

These four cohorts computed specifically year by year will allow the analysis of a combination of different indicators.

(i) Promotion:

E^{2}_{y} = Enrolment in grade 2 in year 'y'

P^{3}_{y+1} = Promotions to grade 3 in year y+1



(ii) Output:

The different rates of output:

(a) by the ratio of the total cycle enrolment which includes intake/promoters plus repreaters in the grade.

(b) by the ratio of the new intake entering the cycle.

(c) by the ratio of completed student years.

(iii) The drop-out rate over several consecutive years.

Therefore, one has the full flow analysis: Enrolment
Intake
Promoters
Repeaters
Drop-outs

in absolute terms and in proportions over a number of specific years.

Such analyses and computations allow many different comparisons to be undertaken according to what unit of analysis is under consideration. From the Bank's point of view, with respect to objectives concerning the improvement of repeater and drop-out rates, the above methods will make possible comparisons of the specific flows in Bank schools with the rest of the schools in the country. The relative temporal changes in the different rates will show whether there is an improvement since the year before, and in relation to other schools.

Flow analysis

The comparison of two (or more) stock enrolment representations. This kind of comparison will reveal the relative changes over one or several years as compared to the base year. Such an analysis ignores, however, the internal dynamics which have determined the observed change. The time-series analysis of stock statistics does show changes. Differences can be compared by examining the proportions in each set of classifications analyzed above. In this way, it is possible to see whether certain proportions (for example, distribution by level) have shifted in importance, or declined in position relative to the preceding year or years. This is perhaps the crudest measure of change, and it will give a reasonable measure of the basic trends in enrolments.

The flow. It was stated above that the stock of students at any given time results from a complex series of flows of students over time. In this section, the flow of students will be examined in detail by grade. Although the methodology will be of a general nature, it must be realized that all flow analyses can be further broken down (according to one's needs) by more specific criteria, such as level, type, sex, age, origin, etc. However, whatever the classification of the statistics, the method remains the same.

Whatever their direction the flow of students between years will allow one to calculate various indicators of the flow:

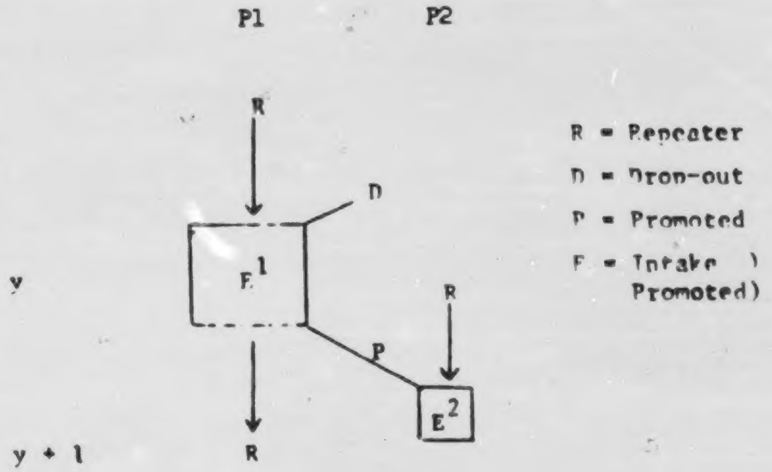
(1) Promotion rate: (i) Specific where the variation of rates at the same level over a number of years will show:
  (a) growth - indicator of better flow potentials
  (b) drop - closing of classes
  (c) static - although growth in enrolments, the same rhythm maintained
 (ii) Average - over a number of years

(2) Repetition rate: (i) Specific
  (a) proportion by year of study or grade to total enrolment
  (b) proportion of repeaters to enrolment in same grade one year ago
  (c) the relation between a and b
 (ii) Average - over a number of years by a and b

(3) Drop-out rate: (i) Specific
  (a) proportion by year of study or grade to total enrolment
 (ii) Average - over a number of years

Graphically, these relationships can be seen in Figure 3-2.

The data required for such an analysis are the grade-by-year distribution of total enrolments, in this case E, and grade-by-year distribution of repeaters. From this information, when two or more years of such data are available, then the number of drop-outs and the new enrolments or promoters are calculated. When such data are available, a full cohort analysis is possible as can be seen in Figure 3-3.


P1
P2
R
D
y
E1
R
P
E2
y + 1
R
R = Repeater
D = Drop-out
P = Promoted
E = Intake)
Promoted)

Grade composition

$$E^n_y = P^{n+1}_{y+1} + R^n_{y+1} + D_y$$

where P^{n+1}_{y+1} = promoted to next grade in next year

R^n_{y+1} = repeaters from same grade in next year

D_y = drop-out of the year of study

E^n_y = enrolment of grade of source

Figure 3-2. Graphic representation of stock and flow of various kinds

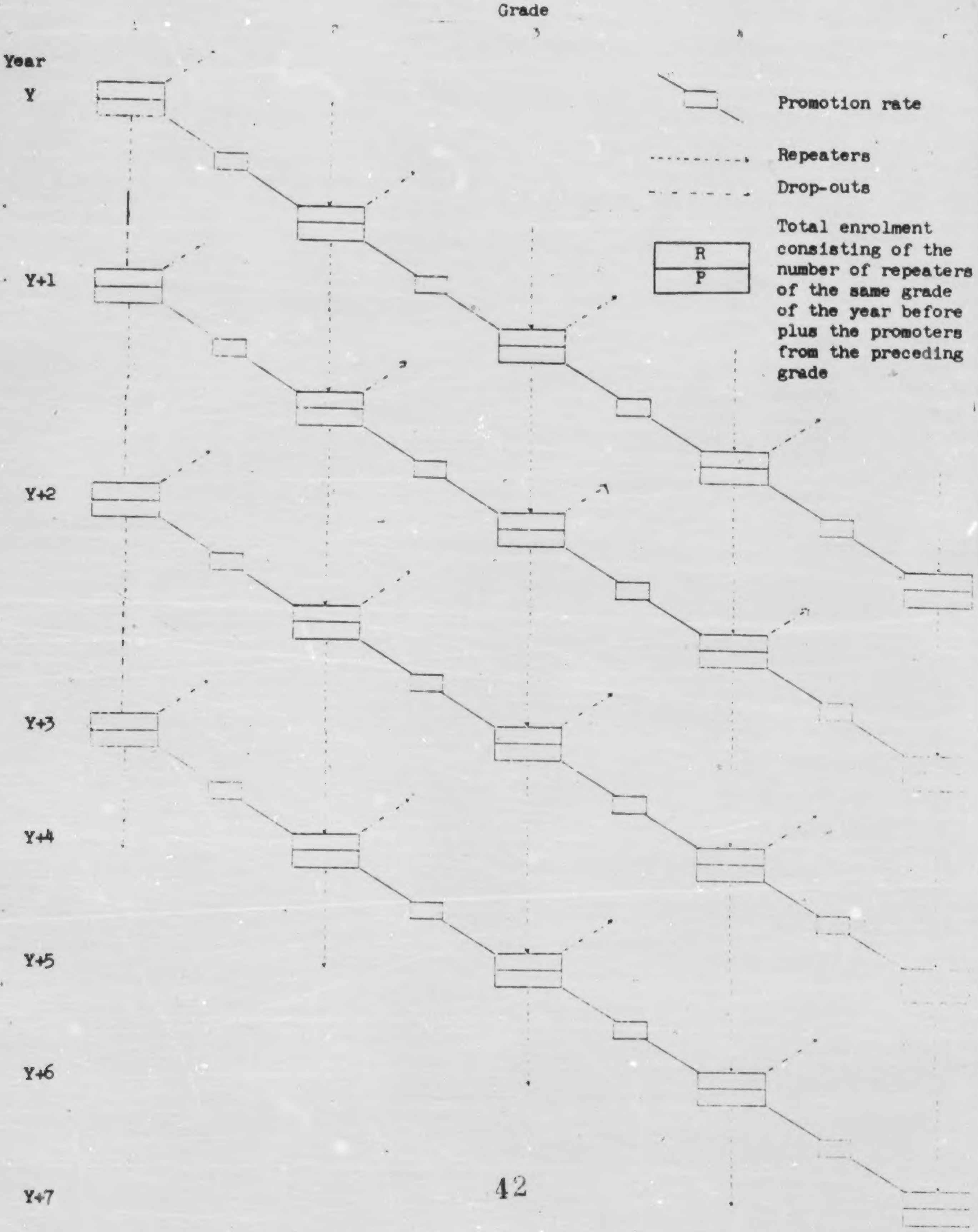

Grade
Year
Y
Y+1
Y+2
Y+3
Y+4
Y+5
Y+6
Y+7
Promotion rate
Repeaters
Drop-outs
R
P
Total enrolment consisting of the number of repeaters of the same grade of the year before plus the promoters from the preceding grade

Figure 3-3. Diagrammatic representation of the way in which stock and flow frequences can be inter-related by grade level and year

Furthermore, where such information is available, the changes over time of the different indicators of the flow of students through school can be examined by regions showing disparities between rates for different parts of the country.

One further analysis can be undertaken from the four specific flow series shown in Figure 3-3. This is the weighted average of the flows. If the rates of promotion, repetition, and drop-out do not vary excessively, a weighted average of the rates, related to a theoretical intake of say 1,000 students, will give a basis for further projections beyond the available years of experience. Such average weights can be used also to compare different schools and regions. Finally, such flows both specific and typical can be examined in the light of existing demographic and market trends and projections.

CHAPTER 4

THE DEVELOPMENT AND EVALUATION OF THE CURRICULUM

The present chapter deals with the process of curriculum development and evaluation. It introduces the concepts of balance and relevance as criteria for judging the adequacy of an educational programme; it then presents a model for generating and selecting educational objectives. Finally, it describes two functions of curriculum evaluation: the modification of curriculum material and the specification of conditions under which the programme should be expected to function well.

Two major themes can be discerned in existing World Bank objectives in curriculum: the balance between subjects, and the relevance of the curriculum to the needs of students or society. The balance between subjects in many cases may imply changes in emphasis between subjects: more emphasis on science subjects in relation to the traditional arts subjects, and more emphasis on practical subjects in relation to academic subjects. The criterion to be used for balancing the curriculum should be its relevance to certain specified needs. From this follows the implication that once a curriculum has been effectively 'balanced', then the curriculum is relevant to attain certain specified needs. The following statement from a Black Book shows just this: "... introduce into the system a more practical bias which is necessary to prepare secondary level children for the employment channels where they will be most needed." (Nigeria, 1964) In this case, the relevance to needed 'employment channels' would serve as the criterion by which balance would be measured. Yet, apart from such an implication of the link between 'balance' and 'relevance', there are no further indications as to what this statement entails in terms of the content of each subject, the learning materials and instructional procedures, in sum the detailed specifications of how such global objectives were arrived at. It is essential that knowledge of how such objectives are specified operationally be available in order that effective evaluation of balance and relevance can be undertaken. To arrive at such specifications, detailed steps must be undertaken which are the bases of curriculum development.

GENERAL PRINCIPLES OF CURRICULUM DEVELOPMENT

Each country must organize its curriculum work along the lines it regards as most suitable to its special conditions. Some countries will have highly centralized curriculum centres while others will attempt to provide for decentralized curriculum work. Some will place their curriculum work in the ministry of education; others will create national centres located in some relation to the ministry, the institutions of higher education, and the public-supported schools; still others will regard curriculum work as a series of special contract projects to be supported by public and private funds.

However it is organized, each country will have to provide for three functions:

(i) The determination of the specifications for the curriculum. This function includes determining the content and objectives of the curriculum. It requires a series of value decisions based on available evidence in the country, studies of specific social and educational needs in the nation, expert opinion, and a body of theory and research on learning. This is the most fundamental set of problems in curriculum development, and each nation may learn from the experience of other nations with these problems.

(ii) The development of learning material, instructional procedures, and the evaluation of their effectiveness. This function requires highly skilled workers who can create materials and procedures that will serve the functions specified for the curriculum and can evaluate its effectiveness. It also includes the tryout of the materials and procedures with students and teachers working under appropriate conditions to identify possible improvements, limitations, and advantages. To assess the effectiveness of curriculum material, it is necessary to create appropriate evaluation techniques, to employ appropriate sampling procedures and research designs, and to analyze the evidence by appropriate statistical and data processing procedures. Again, each country can profit from the experience of other countries and can contribute materials and ideas to other countries. This stage in curriculum development entails a technology which can be shared by all and which can be improved in a systematic way from year to year.

(iii) The in-service and pre-service training of teachers for curriculum changes. Each nation must find ways of bringing its teachers up-to-date on the new subject matter, methods, and ideas finally used in the new curricula. While this is not properly a part of curriculum development, it is clear that new instructional material and instructional methods cannot be used in the schools until the teachers have been prepared for the changes. Here again, each nation can profit from the experience of other nations.

This section is concerned with the determination of the specifications, and with the development and the evaluation of curriculum material and instructional procedures.

DEVELOPMENT OF THE SPECIFICATIONS

Deciding what is to be learned

If an educational programme is to be planned and if efforts for continued improvement are to be made, it is necessary to have some conception of the programme's goals. These goals become the criteria by which materials are selected, content is outlined, instructional procedures are developed, and tests and examinations are prepared. All aspects of the educational programme are really means to accomplish basic educational purposes. Hence, if we are to study an educational programme systematically and intelligently, we must first be sure as to the educational objectives aimed at.

But how are objectives obtained? Since they are consciously willed goals, that is, ends that are desired by the school staff, are they not simply matters of personal preference of individuals or groups? Is there any way to make a systematic attack upon the problem of what objectives to seek?

It is true that, in the final analysis, objectives are matters of choice, and they must therefore be considered value judgements of those responsible for the schools. A comprehensive philosophy of education is necessary to guide these judgements. In addition, certain kinds of information provide a more intelligent basis for applying the philosophy in making decisions about objectives. If these facts are available to those making decisions, the probability is increased that judgements about objectives will be wise.

In view of the principle that investigations can be made which will provide information and knowledge useful in deciding about objectives, the question then arises of what sources can be used for getting such information. Many considerations enter into the determination of what will be taught at any one level in any one subject area. In Figure 4-1, five categories of such considerations have been indicated. Clearly, the critical forms of these considerations will vary from society to society, but some will be always important. Each curriculum development group must decide on the specific forms of these major considerations and examine them.

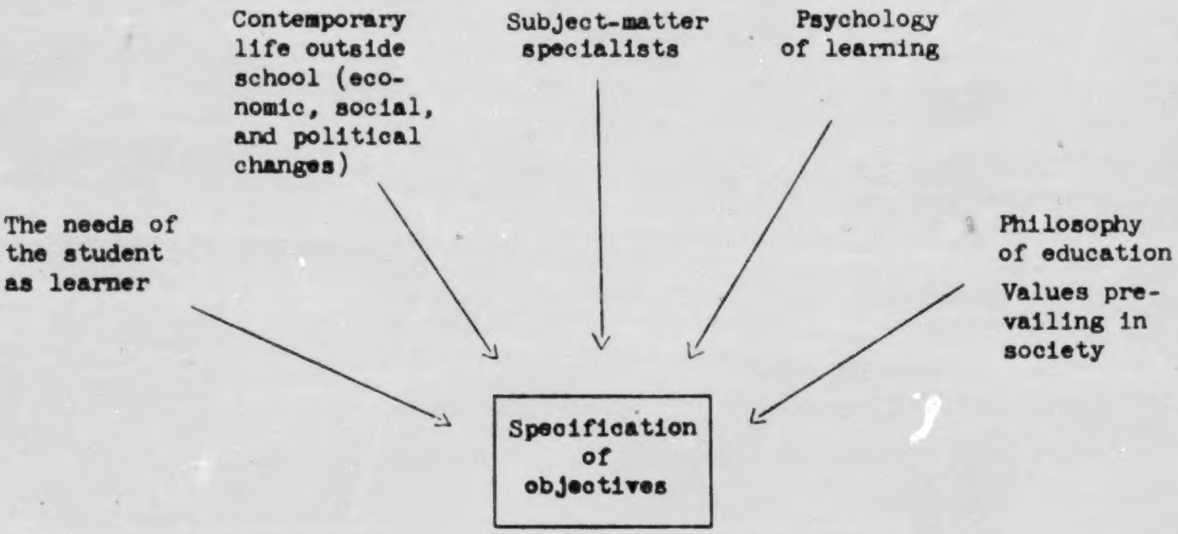

Contemporary life outside school (economic, social, and political changes)
Subject-matter specialists
Psychology of learning
The needs of the student as learner
Philosophy of education
Values prevailing in society
Specification of objectives

Figure 4-1. Categories of considerations in specifying objectives of school learning

Let us examine the five mentioned in turn:

The student as a learner

One categorisation of student needs with which school is supposed to deal could be made according to areas of life such as those presented in Table 4-1.

	Types of skills		
Areas of life	Intellectual	Emotional	Psychomotor
Health			
Occupation			
Social life			
Family life			
Recreation			

Table 4-1. Major areas of life for which the school has to prepare the student. In most of these areas, the student has to acquire some skills which pertain to the intellectual, emotional and psychomotor domains of behaviour though the balance of needed skills pertaining to these three domains is not the same in various areas of life.

In each area of life, studies can be conducted to discover the extent to which all students or sub-groups of students have their needs met. But studies of the learner will generate specific educational objectives only when the information about the present status of the learner is compared with some 'desirable standards', or some conception of acceptable norms, so that the difference between the present condition of the learner and the acceptable norm can be determined. This difference is the identified further need.

Another aspect of the student as learner consists of the relevance to himself or the contemporary world that he sees in what he is learning. Two examples will suffice to clarify the point. Why should a student in an ex-colony continue to learn the history (or geography) of the former colonial power instead of his own country's history or geography? If, in a particular society, drug-taking becomes a problem, to what extent is a special curriculum produced and how quickly is it introduced into the schools?

Contemporary life outside the school

The complex and changing nature of contemporary life suggests that it is necessary to identify the demands that modern life makes upon young people and to equip the young people to meet these demands and the opportunities they provide for development and expression.

What social and economic changes have occurred in recent years? What are the trends? How should these changes and trends influence what is taught in school? If new attitudes about hygiene have been developed through health laws, or if family planning is becoming accepted, how are these introduced into the curriculum? At what levels? As new subjects or in association with what traditional subjects? If the main source of employment is now tourism as opposed to sugar cane production, how should this change influence the curriculum? Needed knowledge and skills can be identified by identifying employment changes and predicted changes. The results of such studies of the types of knowledge and skills required can then be introduced into the curriculum at various points.

Suggestions about objectives from subject-matter specialists

School textbooks are usually written by subject-matter specialists and, hence, their impact has been great. Indeed, over 90 per cent of learning materials have, in most countries, been produced only by subject-matter specialists.

	BEHAVIOR*																			
	A.0 Knowledge and comprehension											B.0 Processes of scientific inquiry I: Observing and measuring					C.0 Processes of scientific inquiry II: Seeing a problem and seeking ways to solve it			
	Knowledge of specific facts	Knowledge of scientific terminology	Knowledge of concepts of science	Knowledge of conventions	Knowledge of trends and sequences	Knowledge of classifications, categories and criteria	Knowledge of scientific techniques and procedures	Knowledge of scientific principles and laws	Knowledge of theories or major conceptual schemes	Identification of knowledge in a new context	Translation of knowledge from one symbolic form to another	Observation of objects and phenomena	Description of observations using appropriate language	Measurement of objects and changes	Selection of appropriate measuring instruments	Estimation of measurements and recognition of limits in accuracy	Recognition of a problem	Formulation of a working hypothesis	Selection of suitable tests of a hypothesis	Design of appropriate procedures for performing experiments
CONTENT	A.1	A.2	A.3	A.4	A.5	A.6	A.7	A.8	A.9	A.10	A.11	B.1	B.2	B.3	B.4	B.5	C.1	C.2	C.3	C.4
1.0 Biological sciences																				
1.1 Biology of the cell																				
1.11 Cell structure and function																				
1.12 Transport of Cellular Material																				
1.13 Cell Metabolism																				
1.14 Photosynthesis																				
1.15 Cell Responses																				
1.16 Concept of the Gene																				
1.2 Biology of the organism																				
1.21 Diversity of Life																				
1.22 Metabolism in Organisms																				
1.23 Regulation in Organisms																				
1.24 Coordination and Behavior																				
1.25 Reproduction and Development																				
1.26 Human Biology																				
1.3 Biology of populations																				
1.31 Natural Environment																				
1.32 Cycles in Nature																				
1.33 Natural Groups and Their Segregation																				
1.34 Population Genetics																				
1.35 Evolution																				
2.0 Physical sciences																				
2.1 Chemistry																				
2.11 Chemical Materials																				
2.12 Classification of Chemical Elements																				
2.13 Chemical Change																				
2.14 Chemical Laws																				
2.15 Energy Relationships and Equilibrium in Chemical Systems																				
2.16 Electrochemistry																				
2.17 Atomic and Molecular Structure																				
2.18 Introductory Organic Chemistry																				
2.19 Chemistry of Life Processes																				
2.110 Nuclear Chemistry																				
2.2 Physics																				
2.21 Kinematics																				
2.22 Dynamics																				
2.23 Energy and Its Conservation																				
2.24 Mechanical Advantage																				
2.25 Mechanics of Fluids																				
2.26 Heat and Kinetic Theory																				
2.27 Wave Phenomena																				
2.28 Sound																				
2.29 Light and Spectra																				
2.210 Static and Current Electricity																				
2.211 Magnetism and Electromagnetism																				
2.212 Electronics																				
2.213 Properties and Structure of Matter																				
2.214 Theoretical Physics																				
2.3 Earth and space sciences																				
2.31 Solar System																				
2.32 Stellar Systems																				
2.33 Meteorology																				
2.34 Physical Geology																				
2.35 Historical Geology																				
2.36 Geophysics and Geochemistry																				
2.37 Oceanography																				
3.0 General																				
3.1 Historical development																				
3.2 Nature and structure of science																				
3.3 Nature of scientific inquiry																				
3.4 Biographies of scientists																				
3.5 Measurement																				

* Some of the student behaviors are given here in abbreviated form. For complete wording, especially of behaviors A.10, A.11, B.4, B.5, and C.1 to C.4, see pages 585-586.

Figure 4-2. Table of specifications for science education

BEHAVIOR*

D.0 Processes of scientific inquiry III: Interpreting data and formulating generalizations						E.0 Processes of scientific inquiry IV: Building, testing, and revising a theoretical model						F.0 Application of scientific knowledge and methods			G.0 Manual skills		H.0 Attitudes and interests						I.0 Orientation				
Processing of experimental data	Presentation of data in the form of functional relationships	Interpretation of experimental data and observations	Extrapolation and interpolation	Evaluation of a hypothesis under test in the light of data obtained	Formulation of generalizations warranted by relationships found	Recognition of the need for a theoretical model	Formulation of a theoretical model to accommodate knowledge	Specification of relationships satisfied by a model	Deduction of new hypotheses from a theoretical model	Interpretation and evaluation of tests of a model	Formulation of a revised, refined, or extended model	Application to new problems in the same field of science	Application to new problems in a different field of science	Application to problems outside of science (including technology)	Development of skills in using common laboratory equipment	Performance of common laboratory techniques with care and safety	Manifestation of favorable attitudes toward science and scientists	Acceptance of scientific inquiry as a way of thought	Adoption of "scientific attitudes"	Enjoyment of science learning experiences	Development of interests in science and science-related activities	Development of interest in pursuing a career in science	Relationships among various types of statements in science	Recognition of the philosophical limitations and influence of scientific inquiry	Historical perspective; recognition of the background of science	Realization of the relationships among science, technology, and economics	Awareness of the social and moral implications of scientific inquiry and its results
D1	D2	D3	D4	D5	D6	E1	E2	E3	E4	E5	E6	F1	F2	F3	G1	G2	H1	H2	H3	H4	H5	H6	I1	I2	I3	I4	I5

The use of subject-matter specialists has been criticized on the grounds that the objectives they propose are too technical, too specialized, or in other ways inappropriate for many students. Probably the inadequacy of many lists of objectives suggested by subject-matter specialists grows out of the fact that these specialists have not been asked the right questions. Often they have thought they were answering the question: What should be the elementary instruction for students who are later to carry on much more advanced work in the field? Hence, a report in history, for example, seems to present objectives for the beginning courses for persons who are training to become historians. Similarly, a report in mathematics outlines objectives for the beginning courses in the training of a mathematician. This is obviously not the form of the question that subject-matter specialists should be asked regarding the secondary school curriculum.

The question which the subject-matter specialists should be asked runs somewhat like this: What can your subject contribute to the education of young people who are not going to be specialists in your field? What can your subject contribute to the layman? If subject-matter specialists can present answers to this question, they can make an important contribution because, presumably, they have a considerable knowledge of the specialized field. Many of them have had an opportunity to see what the subject has done for them and for those with whom they work. Knowing the field as well as they do, they ought to be able to suggest possible contributions that their discipline and its content might make to others.

Quite often the subject-matter specialists specify the requirements of a course in terms of content units to be taught. Thus the historian will indicate that the student should learn about Greek democracy or the French Revolution, or the specialist in Physics will say the student has to learn Kinetics, Dynamics, etc. It should be noted that the list of content itself does not provide sufficient basis for formulating educational objectives. One has to indicate the specific skills or behaviour patterns which the student is expected to apply to each content unit. In the cognitive domain, skills like Knowledge, Comprehension, etc., were identified. In the affective domain behaviour pattern like interest, commitment, etc., were identified. Educational objectives are sufficiently specified only if their definition contains both the specific (mental) behaviour and the unique content area to which this behaviour should be applied. Two-dimensional grids

Enter 0, 1 or 2 in each section of each box.

0 indicates that this topic, interpreted as described in the explanatory notes for the population in question, is not normally studied with this particular objective in mind at this sampling level.

1 indicates that this topic is included in the Science curriculum at this level with this particular objective in mind but is regarded as of slight to moderate importance.

2 indicates that this topic is included in the Science curriculum at this level with this particular objective in mind and is regarded as of major importance.

Subject areas	Objectives							
	1	2	3	4	5	6	7	8
	Obtaining Scientific Information	Interpreting Scientific Information	Theorization Construction	Theorization Utilization	Comprehension	Application of Scientific Knowledge	Personal and Social Objectives	Philosophical Aspects
Earth Sciences								
1. Solar system								
2. Stellar systems								
3. Meteorology								
4. Earth's crust								
5. Physical Geography and Geology								
6. Soil Sciences								
Biological Sciences								
7. Cell structure and function								
8. Homeostasis								
9. Cell metabolism								
10. Cell responses								
11. Concept of the gene								
12. Diversity of life								
13. Metabolism in organisms								
14. Regulation in organisms								
15. Co-ordination and behavior								
16. Reproduction and development								
17. Human Biology								
18. Natural environment								
19. Cycles in nature								
20. Natural groups and their segregation								
21. Population genetics								
22. Evolution								
Physical Sciences:								
Chemistry								
23. Chemical change								
24. Electro-Chemistry								
25. Chemical laws								
26. Chemical processes								
27. Periodic system								
28. Energy relationships in chemical systems								
29. Rate of reaction								
30. Raw materials and chemical substances								
31. Chemical structure								
32. Polymerization and polymers								
33. Chemistry of life processes								
34. Nuclear Chemistry								
Physics								
35. Measurement								
36. Time and movement								
37. Forces								
38. Dynamics								
39. Energy and machines								
40. Mechanics of fluids								
41. Introductory heat								
42. Change of state								
43. Kinetic theory								
44. Light								
45. Vibration and sound								
46. Wave phenomena								
47. Spectra								
48. Static electricity								
49. Current electricity								
50. Magnetism, electro-magnetism and alternating currents								
51. Electronics								
52. Molecular and Atomic Physics								
53. Theoretical Physics								

Source: Comber, L.C. and J.P. Keeves, Education in nineteen countries, International Studies in Evaluation I, Appendix II, Stockholm, Almqvist & Wiksell, New York, Halsted Press, John Wiley & Sons Inc.

Figure 4-3. Subject area/behavioural objective grid

where one dimension specifies content units and the other dimension represents skills are illustrated in Figures 4-2 and 4-3. Each cell of these figures may specify an objective if the combination of the marginal elements of the table yields desired behaviours.

The psychology of learning

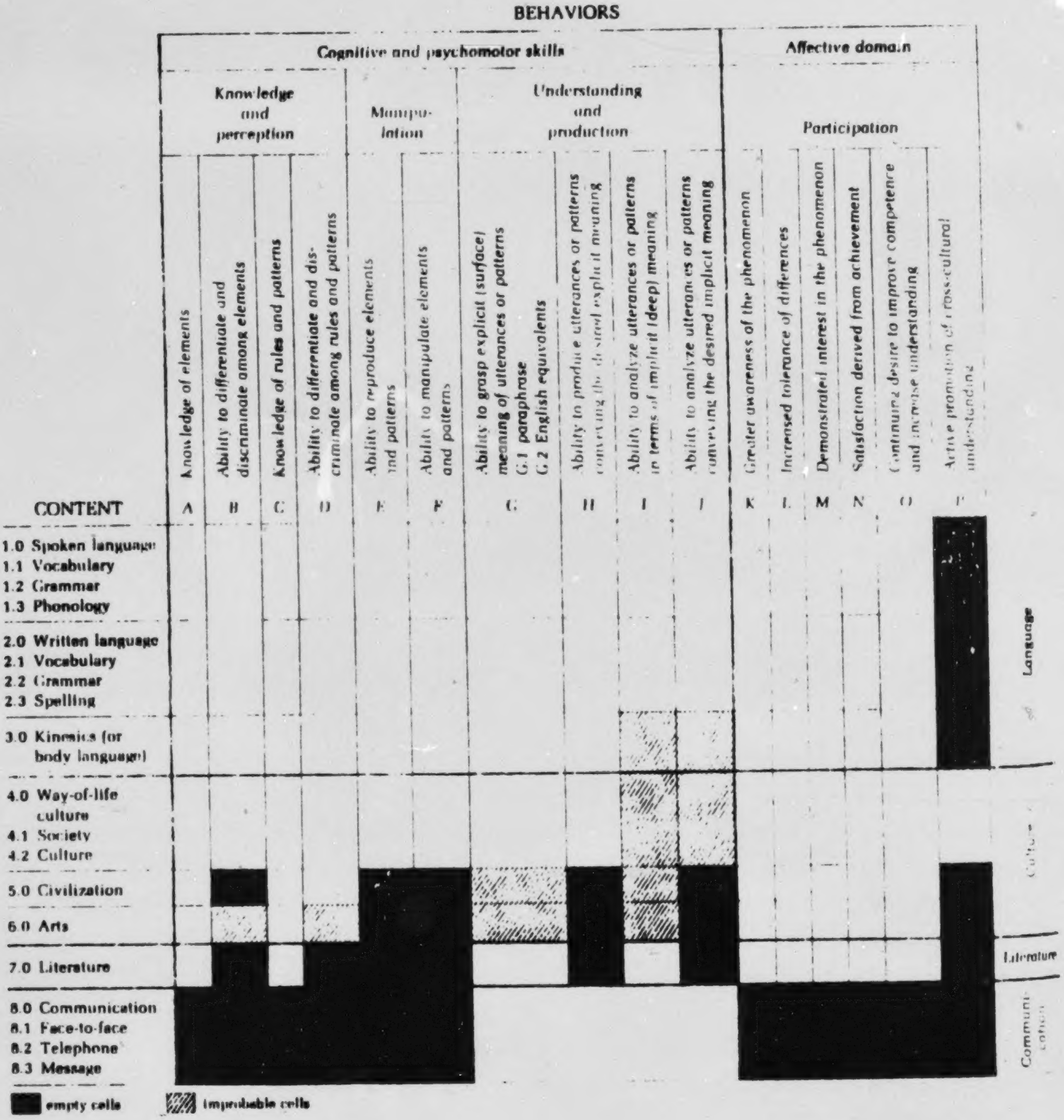
Educational objectives are educational ends; they are results to be achieved from learning. Unless these ends are appropriate to the conditions inherent in the learning process, they are worthless as educational goals.

Thus the psychology of learning does not constitute an independent source of educational objectives; rather it constitutes a set of considerations useful in selecting among goals generated by considering the needs of the learner, the demands of contemporary life, and the topics included in various subject matters. The knowledge accumulated in the field of psychology of learning may help to determine what goals can be attained through school learning, and also to determine the appropriate sequence of goals.

At the least, a knowledge of the psychology of learning enables us to distinguish changes in human beings that can be expected to result from a learning process from those that cannot. Young children can learn to channel their physical reactions in more socially desirable directions as a matter of learning, but it is not possible through learning to inhibit physical reaction altogether.

The psychology of learning can also contribute knowledge as to the time required to bring about certain types of changes in young people. Thus, to change the basic attitudes of children requires continuous emphasis extending over several years. In general, basic attitudes are not markedly shifted by a few months of instruction. In similar fashion, data have been obtained regarding the time involved in changing other types of behaviour, such as ways of thinking and studying, basic habits and practices, interests, and the like. Obviously, psychological knowledge of this sort is useful in suggesting the length of time over which particular objectives will need to be emphasized.

The psychology of learning can also contribute knowledge about stages of cognitive development. Knowledge in this field can determine what objectives are most likely to be attainable at various age levels.



BEHAVIORS																	
	Cognitive and psychomotor skills										Affective domain						
	Knowledge and perception				Manipulation		Understanding and production				Participation						
	Knowledge of elements	Ability to differentiate and discriminate among elements	Knowledge of rules and patterns	Ability to differentiate and discriminate among rules and patterns	Ability to reproduce elements and patterns	Ability to manipulate elements and patterns	Ability to grasp explicit (surface) meaning of utterances or patterns G.1 paraphrase G.2 English equivalents	Ability to produce utterances or patterns conveying the desired explicit meaning	Ability to analyze utterances or patterns in terms of implicit (deep) meaning	Ability to analyze utterances or patterns conveying the desired implicit meaning	Greater awareness of the phenomenon	Increased tolerance of differences	Demonstrated interest in the phenomenon	Satisfaction derived from achievement	Continuing desire to improve competence and increase understanding	Active promotion of cross-cultural understanding	
CONTENT	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1.0 Spoken language 1.1 Vocabulary 1.2 Grammar 1.3 Phonology																■	Language
2.0 Written language 2.1 Vocabulary 2.2 Grammar 2.3 Spelling																■	
3.0 Kinesics (or body language)									▨	▨						■	
4.0 Way-of-life culture 4.1 Society 4.2 Culture									▨	▨							Culture
5.0 Civilization		■			■	■	▨	■	▨	■						■	
6.0 Arts		▨		▨	■	■	▨	■	▨	■						■	
7.0 Literature		■		■	■	■		■		■						■	Literature
8.0 Communication 8.1 Face-to-face 8.2 Telephone 8.3 Message	■	■	■	■	■	■					■	■	■	■	■	■	Communication

■ empty cells ▨ improbable cells

Figure 4-4. Table of specifications for second language instruction

Values prevailing in the society

Another set of considerations in selecting educational objectives arises from the values prevailing in the society. Values must be considered whenever one set of suggestions regarding objectives proves to be incompatible with others or whenever all suggestions cannot be followed because of time constraints. It is, therefore, necessary to choose a few consistent objectives and determine their priorities.

Such choices require value judgements based on the educational and social philosophy to which the choice-makers are committed. An adequate formulation of this educational and social philosophy provides the set of values education should seek to attain. These values serve as criteria for every proposed objective, and the curriculum builder should check whether the objective is consistent with, in opposition to, or unrelated to these values. Thus, some proposals will be accepted, and others rejected. Then priorities will be set among the accepted objectives.

Some of the values on which decisions must be based might be those that deal with such issues as the following:

Transfer of traditional values	vs.	demands of contemporary life
General education of the citizen	vs.	training of specialists
Spiritual values	vs.	material values
Personal satisfaction	vs.	social success
Acceptance of given social order by the educated	vs.	'revolutionary mission' of the school

Another example of the function of social values in the selection of educational objectives may be a case where changes occurred in the proportion of student population attending school. For instance, a school system may have been receiving about 80 per cent of an age group into its primary schools, and only 5 per cent have proceeded to secondary school. Suppose a decision is made that 100 per cent of an age group will have six years of compulsory primary education, but still only 5 per cent will continue to secondary school. The primary schooling then becomes an end in itself. The new situation raises the question of what objectives every child should master by the end of the six years.

Bringing the considerations together

It must be emphasized that the categories of considerations are not arbitrarily chosen ones but probably have universal applicability. Each country must decide on the weights to be assigned to these considerations for its society or various sub-societies. In the last analysis, a set of value judgements will be made in bringing together the various sets of considerations to determine the specifications of objectives.

Let us return now to the concepts of balance and relevancy. How does the foregoing approach ensure that these concepts will be served? Typically, a group of curriculum developers will decide on the subjects to be taught at any one grade level. This is a necessary first step. But as each segment of the curriculum is subjected to the type of systematic development described above, it may turn out that certain parts require more time and other parts less time. Considering this fact and devoting appropriate time in the school programme to each educational objective, one can arrange an 'appropriate' balance. But it should be kept in mind that balance is a time-bound concept. What may constitute a balance at one time may not be appropriate five years later. Changing conditions may call for an increase or a decrease in the amount of attention paid in the school programme to, say, business education, wood or metal shopwork, or home economics. To keep abreast of social change and to ensure balance and relevance to changing conditions, periodic re-examinations of the total curriculum of any educational system are necessary.

Variation between countries

There may be variation among countries in the extent to which they wish to undertake such an approach to curriculum development. There may even be comment on the extent to which such an exercise is desirable. It would seem clear, however, that overall systematic attacks are superior to ad hoc curriculum development (e.g., curriculum development by inspectorial groups 'dreaming up' new curricula). As has already been mentioned, the considerations to be taken into account will vary in the weight they receive from situation to situation but in each case, one should be able to list the forces. A systematic approach is needed especially in cases when alternative curricula are to be developed. Only by identifying different groups of students, such as different school types or different ethnic groups, and systematically determining the curricula for each can one justify commonalities and differences in the programmes.

THE WRITING AND EVALUATION OF MATERIALS

Once the subjects or various domains of learning have been identified for each grade, the learning materials must be produced. These usually take the form of textbooks but often include various audio and visual (non-textbook) components. Indeed, television is being suggested as a main medium of education for some developing countries (but it should, perhaps, be added here that the formative evaluation of television learning units has rarely, if ever, been carried out).

The materials have to be written to the specifications. Perhaps the most important requirement at the materials-writing stage is that there be agreement by panels of writers as to what each writer will produce. A special skill is required for writing materials - a skill which takes some developing. In general, experienced teachers form the best panels of writers. They know the type of language to use for particular groups of children, they learn from each other's writing, and they can use the results of the various formative evaluations. In particular, they know what activities or skills teachers are prepared or not prepared to perform as part of regular teaching and, therefore, they know best how to integrate teaching methods or situations with the learning materials.

Evaluation of the learning materials

Thus far the importance of a systematic approach to the selection of educational objectives and the writing of curriculum material has been emphasized. Systematic procedures are prerequisites for the production of a satisfactory curriculum, but they are not sufficient conditions for it. No matter how systematically the work in the preparation of a programme has been done, and however high the level of competence of the persons who have written the programme, systematic evaluation of the programme is still needed. The evaluation can produce suggestions bearing on two different issues: (a) the modification of the curriculum material, and (b) the specification of conditions under which the programme should be expected to function well.

Modification of the curriculum material. Suggestions for modifying curriculum material can be obtained through utilizing expert judgement immediately after the first draft of the material has been prepared and through collecting data of various types during the preliminary tryout of the material on a small and representative sample of students. Suggested modifications may consist of omitting some portion of the programme, providing more exercises for ensuring mastery of some skill, simplifying the presentation of some idea, providing more visual or manipulative aids to learning some new concept or skill, etc. Generally, it has proved to be useful to submit any newly developed curriculum material (e.g., textbooks, film loops, television programmes, or audio-visual aids) to experts who did not participate in the development and to invite them to judge the material as to its merits and shortcomings. Experts of different kinds can contribute in this way. The subject-matter specialist can examine the validity of the material. The curriculum expert can analyze the material as to whether it contains suitable learning activities. The most valuable judgements can be obtained from experienced and well-trained teachers who are familiar with the goals of the newly developed programme. Such teachers can made judgements as to which learning activities are likely to be smoothly carried out in the classroom, and which learning activities may cause difficulties because of their complexity, vagueness, or lack of appeal to the students.

While such an exercise of judging curriculum material may seem to be obviously necessary, it is far from being universally employed. Systematic judgement of curriculum material usually results in suggestions for modifications that improve the programme. The preliminary tryout of new curriculum material on a small sample of students, prior to its full-scale dissemination, also serves to detect flaws in the programme and may also yield useful suggestions for its modification. Such a tryout of each programme should be made even if it extends the time needed for the preparation of a new educational programme.

During the tryout, data on the success of the programme can be obtained from the following sources:

- Teacher conferences devoted to discussion of the merits and flaws of the programme.
- Teachers' responses to questionnaires prepared especially for this purpose.
- Observation of the teaching process by teams of evaluators.
- Analysis of students' products, such as their responses to questions on work sheets, their preparation of exhibits, etc.
- Students' responses to examination questions prepared especially for the purpose of assessing the outcomes of specific learning activities.

Appropriate data of these types, and their skilful summarization are likely to result in a series of suggestions for modification of the curriculum material.

Specification of conditions under which the programme is expected to function well. Another type of result from the evaluation of new study material is the specification of conditions under which the programme can be expected to function well. The conditions may relate to the availability of some required equipment, teacher competencies, or student characteristics.

An example of such student characteristics can be obtained if we assume that a grade 8 course in physics has been written. Assume also that in a country there is an ability range of students in grade 8 which is highly correlated with the urban-rural distribution of occupations. Here then are two variables which are thought to be important: student ability, and the rural or urban location of school. If the physics course is meant to 'work' for all students in grade 8 in the country (one can, of course, define the particular target population in any way one wants, but this must be done before the whole development process begins), then a small sample based on judgement of classes must be chosen. Figure 4-5 represents a possible sample of classes.

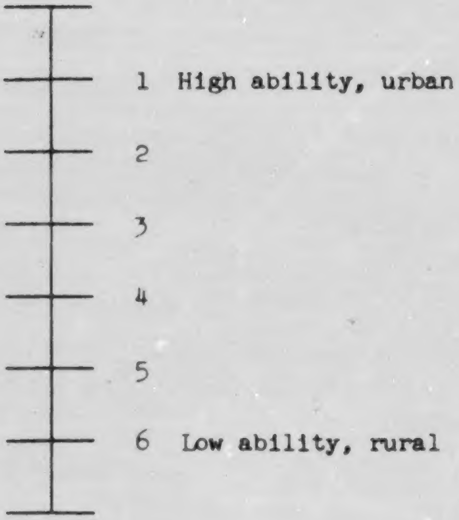
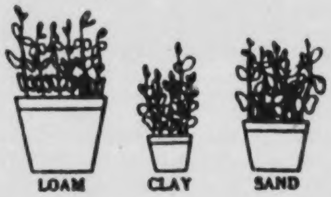


Figure 4-5. Selection of classes across ability ranges

The actual variables used to select schools will vary from society to society, and it is up to the curriculum developers to decide which are critical. The teachers from these classes are trained by the curriculum developers in the utilization of the special curriculum in physics, together with apparatus needed, particular methods of presentation, etc. The instruction then takes place. Throughout the instruction a set of checks and evaluation are conducted. Since the curriculum is meant to be mastered by at least 80 per cent of all students, each teacher keeps a record of difficulties of parts of the course for all students or some of them. At the same time the school principal reports any difficulties he notices in terms of any disruption to the normal working of the school. But the most important criterion of how well the curriculum is 'working' is the learning of the students. Since the specific objectives of each part of the new curriculum are known, it is possible to determine the extent to which the children have mastered a particular example. If one objective was to understand the principle of experimental design, an item such as the following might be used to test whether such understanding had occurred:

> Tom wanted to learn which of three types of soil - clay, sand, or loam - would be best for growing beans. He found three flowerpots, put a different type of soil in each pot, and planted the same number of beans in each, as shown in the drawing. He placed them side by side on the window sill and gave each pot the same amount of water.



> Why was Tom's experiment NOT a good one for his purpose?
>
> A. The plants in one pot got more sunlight than the plants in the other pots.
> * B. The amount of soil in each pot was not the same.
> C. One pot should have been placed in the dark.
> D. Tom should have used different amounts of water.
> E. It would get too hot on the window sill.

Let us hypothetically assume the following percentages of students succeeding:

Total	Urban	Rural	School
75	81	71	1,2,3

These figures would indicate that:

(a) The learning materials are not working well enough for the target population for this objective since we decided that at least 80 per cent should master it.

(b) The materials do not fit the needs of rural children but they are appropriate for urban children.

Thus it can be said that use of the programme should be limited to schools in urban areas, or in other words, one condition of using the programme has been specified.

IMPLICATIONS FOR WORLD BANK

This section will deal with the World Bank's possible actions concerning the development and evaluation of curricula.

Elaboration of the specifications in World Bank objectives in curriculum

The selected objectives stated at the beginning of this chapter are vague and although indicative of the intention, require a great deal of elaboration. (The statements are rather analogous to stating "A new school will be built" without indicating where, for what purposes, for what children, at what cost, etc.) A selection of qualifiers from the following list should be added:

(i) what subject areas?
(ii) what grade levels? and target populations (e.g., type of school or region)?
(iii) what considerations are to be taken into account?
(iv) what personnel (with what experience and qualifications) will develop the curriculum?
(v) over what time period?
(vi) at what cost?
(vii) with what forms of evaluation (e.g., forms based on children's responses, teachers' views, or principals' views)?

If one returns to the statement in the Spain 1972 Appraisal Report "Development of new curricula for technical, vocational and agricultural education", one might think of a statement as follows:

> "A specially convened group of persons within the Ministry of Education will systematically develop new curricula for Grades __ to __ for ___________ school types in the subjects of technical, vocational and agricultural education. This group will examine economic and social changes within the society as well as changes in the subject matter itself in order to produce the learning objectives accompanied by learning materials. The materials will be subjected to formative evaluation, the aim being that 90 per cent of the students in the target group reach mastery of all objectives. It is anticipated that this curriculum will be developed to the point of implementation within 30 months of the inception of the project."

World Bank evaluation of work undertaken

How can the success of a project with these kinds of objectives be evaluated by the World Bank? A first approach would be that of interviewing curriculum and general education personnel in each country: curriculum developers, teachers, school administrators, and educational advisers. The interviewers would, themselves, need to be knowledgeable about both the principles of curriculum development and the culture of the country.

In essence, this involves checking that the general principles have been adhered to. The curriculum developers, if they have done their work well, will be only too pleased to inform a 'World Bank person' about what they have done and show the results of various studies and formative evaluation. The 'World Bank evaluator' will then be in a position to know how well the work has been conducted.

The real problem arises in establishing the curriculum group in each country. If the group is good, the work will flow well. Further attention will be paid to the establishment (and training) of such groups in Chapter 7.

EVALUATION OF CURRICULUM IMPLEMENTATION

Several World Bank statements assume that revised or improved curricula have already been developed and that the main task is now to implement such curricula. From an experience, it is rare for such curricula to have been revised or developed systematically, and it would be desirable for the World Bank to check on such development. This has already been referred to above. Once it has been established that the general principles of curriculum development have been adhered to, then the improved curriculum is ready for implementation.

It is at this point that the actual process of evaluation begins. For it is now necessary to ascertain the extent to which the desired curriculum is in fact being implemented. The main approach used in curriculum evaluation is through the testing of student achievement, and the methods and techniques involved in this are discussed below. Before considering these, however, two other aspects will be briefly discussed.

School timetables. One useful, but limited approach is to take the nationally prescribed timetables (where they exist) both before and after implementation. Comparisons of the subjects offered and the time allocated to each subject at the grade-level(s) in question would be one indication of whether a change had taken place, and the extent to which this change was meeting the desired curriculum. Where nationally prescribed timetables do not exist, the 'World Bank evaluator' could himself draw a 'judgement sample' of schools and ask for the two sets of timetables in order to make the appropriate comparisons.

In either case, the 'World Bank evaluator' should make some on-site visits to schools he has selected, to discover the extent to which the time allocations stipulated are actually being observed, and the space and materials required for the proposed curricular activities are being provided and used. Such time allocations as are examined both before and after implementation will show how much time a student is expected to spend on different parts of the curriculum. The basic limitation of this approach, however, and it is an important one, is that it does not really show how much is being learnt by the students themselves.

The analysis of the returns from the testing process

The results from the various types of testing that may have been undertaken are now fed back to the curriculum development team for scrutiny and analysis. To illustrate such analysis suppose that the curriculum planners are looking for a gain in performance on all topics represented by a simple item in the test. When the curriculum was tried out such a gain was actually obtained, but it is not possible to tell if this is going to be so after it has been implemented over the whole country. The returns from the evaluation procedures show that in some cases a gain has been obtained to a greater or lesser degree, but in certain schools or in certain sub-groups of the population there seems to be no increase at all, and in some cases a fall in performance. Such cases have to be examined in detail in order to determine what are the problems that are hindering the efficient working of the curriculum.

So far we have been talking about one type of testing of the curriculum - the use of multiple-choice items. Multiple-choice items will give an immediate idea of the number or proportion within each chosen sub-group (according to critical variables and sampling strata used) who have grasped (or not) a particular subject area in a particular behavioural category. However it is important that certain limitations of multiple-choice items be discussed in order to broaden the range of the type of evaluations undertaken in curriculum. The question to be posed at the moment is: are multiple-choice answers necessarily the only or surest way of evaluating curriculum? There are other written tests, performance tests and oral tests, that may give other indications of achievement in the new curriculum. Performance tests are important in subject fields like science, agriculture, arts and crafts, etc. By systematic observation it is possible to determine to what degree have students acquired skills using their hands or using tools. The information obtained from written tests and from performance tests can be supplemented by observing actual behaviour of teachers and students in the class. It is advisable that the observers should assess activities on the basis of a check-list of desired events which are to occur in the class. Observation enables to determine to what degree exists correspondence between planned and actual curriculum activities.

It has sometimes been expressed by those concerned with development of any curriculum in developing countries that positive attitudes towards manual work should be a fundamental part of the instructional process. The introduction of practical or relevant subjects into curriculum is certainly a necessary and important first step, but paper and pencil tests like the above do not give any indication of the present attitudes or change in attitudes.

All this leads one to conclude that any evaluation of student achievement should be, time and money permitting, undertaken using both written and performance tests, and observations of actual student behaviours.

The place of national examinations in the evaluation process

Nationally devised external examinations are set at one or more stages in the education system in most countries, and it is often thought that the performance of students on such examinations can be used to evaluate the curriculum they have been following. It perhaps needs to be stressed that this is not possible for reasons which will be made clear in the rest of this section. It should be noted, however, that examinations are usually devised to serve two major purposes: (a) to assess how much students have learned in a particular course or programme, and (b) to enable predictions to be made about possible performance on some future course or programme. Examinations for both these purposes may be set internally within separate schools or they may be set externally by the State or some central authority. We are only concerned here with the latter, where examinations for the first purpose may be used, for example, to issue School Leaving Certificates, and for the second purpose to select those students suitable for moving on to a higher stage of education.

Our first concern is with external examinations that are supposedly set to yield a measure of how much students have learned. The point we wish to emphasize is that in all such examinations, their actual type and content exerts a profound influence upon what is taught in the schools in the year preceding the setting of the examination. This 'backwash' effect is not confined to developing countries; it also occurs in many more advanced countries. We should like to put on record the judgement that examinations can be good servants of an education system, but should never become its master. In other words, the curriculum should dictate the content and objectives of examinations, never the other way round. An end-of-course external examination is a very powerful instrument which can be badly misused; for it is an unsound educational principle to allow an examination to determine what students need to learn and hence what they will be taught.

There is an ideal sequence governing the development of a specific course or unit of instruction, in which the first step is the production of a set of detailed behavioural objectives. This is then used both as a guide in the devising of those instructional procedures which it is hoped will lead to the achievement of the objectives, and as a basis for the 'blueprint' for the production of an examination to assess the extent to which students have achieved the objectives. If this sequence is followed, it can be seen that not only is the course or unit 'tailor made' to fit the desired objectives, but that the examination will also be designed to measure exactly what the students have been taught.

The fact that the ideal is rarely achieved makes it imperative that the examination system employed in every country containing a World Bank project in education should be given close scrutiny, whether or not there are specific objectives related to this area. It may indeed be necessary to seek a reform of the examination system in a given country, simply in order that other educational objectives may stand some chance of being achieved. For example, in many projects there are objectives concerned with the introduction into the curriculum of 'a practical bias'. Clearly what is precisely meant by this statement needs to be detailed in specific terms, but if the countries concerned in some subsequently administered examination introduce no element of 'practical' assessment, the chances are very high that this element will be given low priority in the teaching, resulting in the non-achievement of the objective. It is important that some measurement of manual ability, appropriate to the desired curriculum, be included, in order to ensure that the intended practical bias in the curriculum is taught.

The science of examining is one in which much progress has been made in the past thirty or so years. Despite this, there are too few people in most developing countries with sufficient background, knowledge, training, and experience in the area of examinations to be aware of the problems and difficulties. Without dealing too exhaustively with this subject, it is relevant in this context to mention two important attributes of any examination: its reliability and its validity. The reliability of an examination - that is, the consistency with which it measures whatever it does measure - while fundamental, is sometimes emphasized rather too much. The changeover to objective, multiple-choice tests, for example, while certainly increasing reliability, often does so at the expense of validity which is a far more important attribute. An examination will have high validity if it measures what it was intended to measure, and it can be seen that the ideal development mentioned above makes certain that high validity is 'built in'. An examination development process which ensures that students are measured on the same objectives that they have been taught during their course, also takes care that the backwash effects are beneficial. If a student prepares himself specifically to pass his end-of-course examination, he will be achieving the very objectives that the course set out to teach him. An examination, however, which is constructed independently of the designing of a course, while it may have a reasonably high reliability, may nevertheless possess relatively low validity, and at the same time, by emphasizing matters of less importance, exert a detrimental backwash effect.

We believe that this is an area where more help and guidance could be given and certainly one where assistance in providing the appropriate training of indigenous personnel would be handsomely rewarded. We also eel that, because any curriculum change is necessarily related to (since it is influenced by) the examination system, more projects which are concerned with improving education by introducing changes in curricular methods and approaches, should have incorporated in them aspects directly related to the introduction of changes in the examination systems of the countries concerned.

Our second concern with existing national examinations is that in the vast majority of cases, the same examination is used for both the purposes mentioned above. That is, the same examination will be used for (a) awarding a Leaving Certificate from a primary school, and (b) for selecting those students best suited for advancing to secondary education. A moment's reflection will reveal the impossibility of adequately combining these two functions into one examination. For the first, it is clearly necessary to include, with correct weighting, all aspects of the primary curriculum, whereas, for the second, the best prediction will be obtained by limiting the examination specifically to the more academic subjects.

In practice, in countries where such examinations are given at the end of primary schooling, they are usually confined to academic subjects, with the negative 'backwash' effects referred to above. If, however, the content of the examination was broadened to relate more precisely to the desired curriculum (including, for example, assessments in the areas of agriculture and rural arts), but the examination was also to be used to select students for secondary education, then again, this function of the examination would inevitably predominate and the primary curriculum distorted by a greater emphasis being given in the teaching to the academic subjects known to be important in the secondary school. The point that must be stressed here is that, where both a School Leaving Certificate and a selection process are required, then separate procedures must be employed.

If a formal external examination is retained to assess students for a School Leaving Certificate, the selection aspect could be performed by utilizing a scheme employing teachers' assessments and tests of verbal reasoning. It must be added, however, that to do this, some additional training of the teachers would be required. Alternatively, School Leaving Certificates could be based upon a procedure utilizing some form of continuous assessment. Again it must be added, that such schemes require a fairly high degree of sophistication in assessment procedures which it may take many years to develop in the teachers concerned.

The final point which must be made in this discussion of examinations is that only in rare circumstances can the formal external examination have any part to play in the process of curriculum evaluation. Examinations by their very nature are designed to assess individual students: evaluations are concerned with the curriculum - that is, with the over-all performance of all students following it. As will be shown in the following section, there is no interest in the evaluation process with the performance of individuals. The one possible exception, which will not be elaborated here, is where the results obtained from an external examination which has been carefully designed to assess student performance at the end of a specific course, are also used collectively to assess the curriculum in a summative evaluation.

Systematic evaluation of curriculum through measures of student achievement

There is a close connection between curriculum evaluation and the assessment of achievement, because a change in the curriculum normally implies some change in the achievement of the students. This remains true whether the curricular change considered is the initial introduction of science teaching into primary schools, the modernization of mathematics teaching in the secondary programme, or a field course in modern agricultural techniques for primitive farmers. Therefore the documents that accompany the educational projects supported by the World Bank contain many statements of purpose that imply changes in the level of student achievement. Examples of these are expressed in the following phrases taken from the project documents that we have studied in detail:

> "The project would raise the general standard of secondary education" (El Salvador, 1966).
>
> "The project will help meet the country's urgent need for the improvement of secondary school education" (Ethiopia, 1966).
>
> "Proposed project will assist the government to improve the quality of education" (Iran, 1970).
>
> "The objectives of the proposed project are to improve the quality of educational programs" (Turkey, 1971).

In general, evaluation of the projects in relation to these aims could be accomplished if the appropriate procedures were followed. The curriculum, whether new or old, needs clear definition, and the educational objectives of the curriculum need to be stated. The assessment is made in terms of these objectives - by performance on examinations or by some other means.

Alternative designs of possible evaluative procedures will be considered later. The point being made here is that a clear statement of objectives is a necessary precursor of objective evaluation.

Analysis of all the project documents we have examined suggests that most of the objectives concerned with levels of achievement are stated vaguely in terms of some overall 'standard' or 'quality'. The foregoing examples are not atypical in this respect. Some projects have introduced a degree of clarity in specifying the subjects in which achievements were to be raised. (Usually these subjects are science or science-related practical subjects.) Examples of such statements are:

> "Science teaching will be extended and given greater emphasis in the schools included in the project" (Kenya, 1966).
>
> "The major contribution of this project item however would be to upgrade the quality of engineering education and to produce engineers capable of solving the practical problems of Spanish industrial development" (Spain, 1972).
>
> "...improvement of agricultural education and training" (Philippines, 1973).

We feel however that it will be necessary to state achievement objectives in a much more precise form is some objective evaluation is to be possible. The specification of objectives is an activity that has received a great deal of attention during the last 20 years. One conclusion which has been generally agreed is that the objectives should be stated in behavioural terms; i.e., they should describe in unambiguous terms the precise behaviour of the student that will result from and give evidence of the achievement of the objective. By this criterion the examples of objectives quoted earlier are clearly inadequate.

For the purposes of evaluating achievement in the educational projects of the World Bank, a much greater degree of specificity is recommended. Sample objectives which communicate the type of behaviour required of the student without loading the statement with superfluous detail are:

- The student will be able to correctly identify the effects of specified soil deficiencies on the cultivation of maize.
- The student will be able to convert measures quoted in pounds and ounces into kilogrammes.
- The student will be able to read and comprehend specified instruction manuals written in English.
- When given an arc of a circle the student will be able to measure the radius of the circle and locate its centre.
- The student will be able to measure the pH of a sample of soil.

Objectives stated in these terms permit the construction of suitable test instruments to evaluate their achievement. It must be noted, however, that if an entire curriculum were to be defined in this detail, the mass of behavioural objectives could become overwhelming. This has led to the widespread use of grids as described in the earlier part of this chapter. The intersection of two dimensions, one of content material and one of the level of behaviour required of the student, permits a complete overview of the objectives of a curriculum. The usual practice is then to enter on to the grid, in each cell, a sample of the specific objectives that relate to that particular cell. Test instruments can then be constructed to match these blueprints.

While we believe strongly that the most important contribution of objective evaluation will be achieved only where the curricular objectives are defined early in the project, it does not appear that they could easily be included in the official documents that precede and accompany the signing of an agreement between the World Bank and government of a developing country. Rather, we suggest that the Appraisal Report recommend a specific commitment on the part of the country to mount a concurrent evaluation of its work, and that a detailed specification of the curricular objectives be filed with the Bank within a limited period (perhaps 12 months) from the inception of the project.

Although most of the literature on achievement evaluation methodology, and indeed many of the examples in this report relate to the secondary phase of education, the ideas are equally applicable at the primary stage and in post-secondary education. They cover both formal and informal education, provided always that there is at least an agreed-upon curriculum. They cover both academic and theoretical subjects on the one hand and practical/vocational courses on the other. In fact, all of these have a place in any complete educational system, and we have noted a proper concern with both academic and practical work in the World Bank projects we have reviewed.

There are two ways in which specific objectives can be stated. They could refer to increase, as in the following example:

> "The project will lead to an increase in the number of secondary school students who understand the theory of crop rotation."

They could also be stated in more absolute terms as follows:

> "A result of the project will be that all secondary school students will be able to explain the theory of crop rotation."

Chapter 3 defines these approaches as 'norm-referenced' and 'criterion-referenced'. Which is the most appropriate formulation will vary according to the situation, but it must be recognized that they have very different implications as far as evaluation is concerned.

Measurement of an improvement or an increase implies that some base measures exist. In many cases it may be necessary to establish these at an early stage of the project. This approach would appear to be particularly useful for relatively unstructured programmes designed, for example, to increase literacy, or to make educational facilities available to nomadic sub-groups of the population. In these instances the progress of a project can be measured in terms of observed changes from the base measures.

Criterion-referenced evaluation does not require the construction of a base measure, but it is weaker without one in that it cannot describe quantitatively how much the project has achieved. It is most appropriately used in situations when one can safely infer that change is taking place. To take a specific example, let us consider the introduction of new agricultural courses for farmers in a relatively undeveloped rural area. If the course material is quite new then criterion-referenced objectives and evaluation seem appropriate (e.g., the students will be able to correctly operate a tractor-drawn plough). On the other hand, if the course aims to spread, more widely, knowledge and skills that already exist within the group, then a norm-referenced approach would be more appropriate (e.g., the students' ploughing skills will be increased). Whichever of these approaches is adopted, a great deal of educational benefit can be gained from a system of continuing assessment providing a flow of information back to the developers. The advantages of such formative evaluation procedures have been discussed at length in an earlier section. They are relevant for achievement evaluation, should be conceived well in advance, and should be built into the plan of the project.

This is not to deny that a certain amount of objective evaluation can be carried out even where it has not been included in the original plans. It will still often be possible at the end of a project to obtain suitable data on the achievement of specified objectives. But inferences as to the direct impact of the project can only be based on estimates of the original standards in the absence of contemporary evidence on these standards. Such estimates are not likely to be reliable or accurate as the memories of individual teachers and inspectors are notoriously invalid.

When the detailed specification of objectives has been made, it is clear that evaluation of achievement needs to be based on carefully constructed tests, rather than on the opinions of curriculum experts, school inspectors, or indeed teachers. Unfortunately, good tests are not easy to construct, and they are often specific to a particular situation. It would be unrealistic to think that one well tried test of achievement in secondary school science, for example, would be appropriate for many of the Bank's projects, since the content of the test would depend critically on the detailed teaching objectives of a particular curriculum.

The test, therefore, will have to be constructed on the spot. Furthermore, although the skills and techniques of test construction can be taught, and numerous educational centres around the world offer courses on this topic, it is likely that only a person working inside the particular educational system will know enough about the curriculum to construct suitable measuring instruments. What is required is that staff with the necessary test construction skills work with pedagogical experts within the country concerned to generate an appropriate combination of generalisable expertise and local insight. If educational evaluation is to be built into a project, and if a baseline of achievement is to be established, this collaboration should occur early in the project's execution. It may well prove that the availability of pools of well tried test items in various subject areas, such as those accumulated by the International Association for the Evaluation of Educational Achievement (IEA) can do much to speed the process of test construction, but under no circumstances should such pools be used without reference to particular local conditions and specific curriculum objectives.

The preceding paragraphs have focussed on the specifics of objectives and tests. These are important parts, but only parts, of the complete evaluative study. The design of the study is itself important. Three alternative models will be discussed.

The first is the simplest, but much the least satisfactory. It depends on the exploitation of an existing national system of examinations, such as the Bagrut examination in Israel, the General Certificate of Education examination (GCE) used both in Great Britain and in certain Anglophone countries, the Baccalaureat used in France and some Francophone countries. The major purpose of these examinations is the assessment of the performance of individual students for certification purposes, and to a lesser extent for the purposes of selection for tertiary education. Yet there have been attempts to use the data obtained from such examinations to obtain evaluations of achievement on a national scale. Serious difficulties arise in such an approach:

(a) The examination questions are designed to discriminate between students and may not be appropriate for showing whether particular curricular objectives have been mastered.

(b) The content of the examination is unlikely to cover the complete curriculum in sufficient depth. Usually only those parts that are taught in all the schools are likely to be included.

(c) Although the common practice of allowing students a choice as to which questions they will attempt to answer may appear fair and generous to the students concerned, it denies the researcher the possibility of knowing how the student would perform on those questions that he does not select. Thus it is impossible to make a comprehensive statement about the achievement of any one student, and in consequence one cannot arrive at a comprehensive assessment of a complete group.

Thus, this approach does not seem very satisfactory. It would perhaps be better to use certification examinations for that purpose only.

The second model to be considered is one in which a nationally organised and authorized body of researchers is continually testing random samples of students of different ages with test materials from a wide variety of curricular subjects. The plan is that a complete curricular evaluation of a subject, such as, for example, mathematics, would occur once every two or three years. Provided the basis for choosing the samples has not been altered in the meantime, the change of performance from one occasion to the next would give evaluators a measure of the extent to which the level of achievement had risen. The chief drawback of this approach is that it is extremely expensive in terms of resources. It needs professional staff working over a long period to carry it through, and requires facilities for the frequent administration of achievement tests to randomly drawn national samples. It appears probable that only a country with an already highly developed educational system could afford such a procedure. In a country in which only a minority of the children complete secondary education, it would be an unjustifiable luxury.

For developing countries, a third model would appear to be more appropriate. This would make use of the methods developed by the IEA, though with the limited objective of measuring achievement in one national system. Random samples of students who have been working with a new curriculum would be drawn (as might be additional samples for comparative purposes). These students would then be included in the testing programme throughout the evaluation of the project if the evaluative mode is formative, or towards the end of the project if the mode is summative. As the performance of groups of students rather than individual students is of interest, advantage would be taken of item sampling techniques whereby different students in one class would take several forms of the test. This considerably increases the amount of test material it is possible to use, and thus provides a much more detailed picture of national achievement. The testing programme would be operated in conjunction with questionnaires to the students - collecting data on educational background and related characteristics - and questionnaires to the teacher - to gather information as to what has been taught, etc. The data so accumulated can then be analyzed by means of various multivariate statistical models so as to yield precise descriptions and explanations for varying levels of achievement.

One could envisage a continuing evaluative effort in cases where a country was engaged in a long-term programme to improve a particular part of its educational system, e.g., secondary science teaching. Here it would be worth establishing a small evaluation unit to carry out the necessary test construction, administration, and analysis. We feel that where the Bank is supporting the initial parts of such an educational programme, it might well feel that it is also appropriate to support the setting up of an evaluating and monitoring research unit. A description of a hypothetical exercise in evaluating student performance is given in Chapter 5.

CHAPTER 5

THE USE OF EVALUATION IN WORLD BANK PROJECTS : AN ILLUSTRATION

The purpose of this chapter is to illustrate how the process of evaluation might be carried out in a World Bank Education Project. It must be emphasized that although this illustration deals with one subject - science - in formal education, the principles and stages described apply equally to other subject areas and, with but few modifications,to evaluations in less formal aspects of education. In other words, the differences between the evaluation described here, and that, say, of a course in a basic education school, a farmer training school, a vocational school or a community education centre, will be largely related to the differences in subject matter, and not to the process itself. At the same time, it is understood that there will be a greater emphasis in future World Bank projects on the development of formal educational institutions so that the choice of this particular illustration is probably appropriate.

The distinction between formative and summative evaluation has been made clear in an earlier chapter, and stress has also been laid on the importance of considering evaluation as a process for improving the education process. Since formative evaluation introduces additional stages, the illustration given will be largely of this kind, and emphasis will be placed on those aspects which are aimed to produce improvements rather than simply to act as a check on what has, or has not been accomplished.

The general nature of the evaluation process was described in some detail in Chapter 2. For the purposes of this illustration the three major aspects of the evaluation model have been selected, and the practical issues which arise from them will be covered in some detail. The competencies of the people required to carry out the work, as well as the time required for the different stages of the whole operation will also be covered.

The major aspects of the evaluation model are illustrated in Figure 5-1.

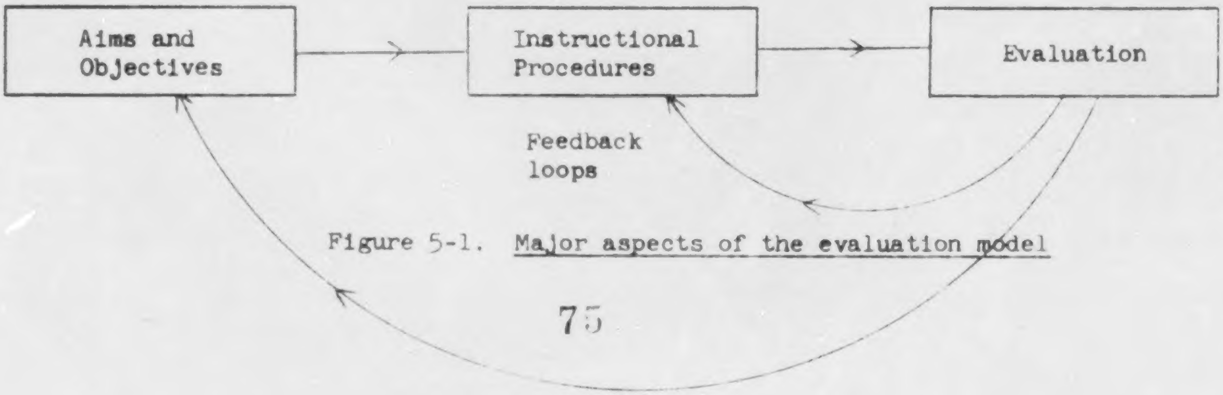


Figure 5-1. Major aspects of the evaluation model

The various aspects to be covered in this illustration are therefore:

(a) Development of Aims and Objectives.

(b) Development of appropriate procedures aimed to achieve the objectives.

(c) Description of the kind of evaluation required, and the methods to be adopted in carrying it out.

(d) Indications of the Time Scale involved.

(e) A note on Summative Evaluation.

DEVELOPMENT OF AIMS AND OBJECTIVES

Mention was made in Chapter 3 of this report to the varying degree of specificity with which objectives may be written. A diagrammatic representation of the three main types of objective(1), according to their degree of specificity is given in Figure 5-2.

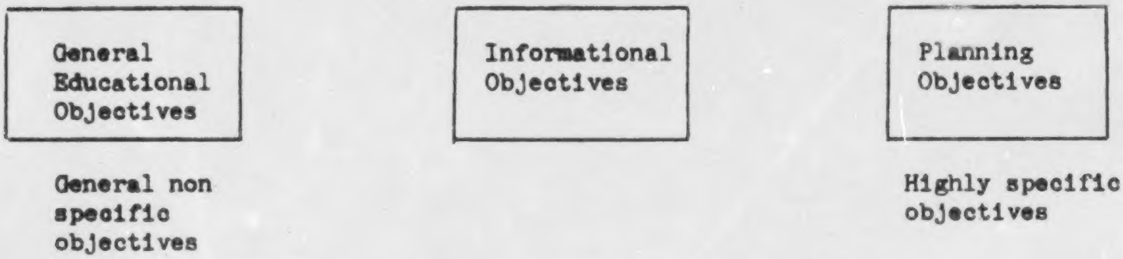

General Educational Objectives
Informational Objectives
Planning Objectives
General non specific objectives
Highly specific objectives

Figure 5-2. Three main types of objective

General objectives, at the left of this scale, are usually expressed as rather vague declarations of intent and, for the most part, correspond to the statements which, in the past, have appeared in World Bank Projects which have not been designed with evaluation in mind. Such general objectives or aims form a necessary part of a project description, but the need for their elaboration becomes evident if an attempt is to be made to develop a new curriculum, or change an old one, which is to be designed to ensure that the intentions of those who have prepared the project are fully carried out. There are, therefore, three stages in the development of objectives:

---

(1) Figures 5-1 ans 5-2 have been taken from Behavioural Objectives and Instruction, G. Kibler, R.J, Basher, L.L, & Miles D.T.Boston, Allyn & Bacon, Inc. (1970)

(i) Writing the general, non-specific objectives declaring the main intentions for which the project is being devised.

(ii) Developing more detailed objectives which will provide the additional information necessary to produce the more specific planning objectives.

(iii) Producing the planning objectives. These will be used (a) to develop the appropriate instructional procedures of the desired curriculum, and (b) to guide the development of the evaluation techniques required.

Each of these three stages will now be described in more detail, illustrated with examples from an hypothetical project. It is not intended to specify the full quantitative details of such a project, but merely to provide sufficient background information to enable the evaluation aspects to be seen in appropriate perspective.

It may be supposed that a project is being considered in a developing country which will provide among other things for the construction of a number of new secondary schools. In addition, it may be specified that the aim is not merely to increase the number of secondary school places, but to introduce a new type 'modern' curriculum in the schools in order to 'improve' the general standard of education. Furthermore, it may also be an intention of the changed curriculum to emphasize particular subjects, such as science, and to introduce new elements such as different aspects of practical work.

Stage (i). Writing general, non-specific objectives

These are the objectives produced at the early stages of the discussions on the new project, and would not be dissimilar from those which appear in the PIM's or Side-letters of existing projects. For example in the example project being considered here the following general objectives might be written:

1. To increase the quantity and improve the quality of graduates from the general secondary schools.
2. To modernize the curricula, teaching methods and textbooks to be used in the schools.
3. To extend the teaching of, and give greater emphasis to, science in the secondary schools.
4. To introduce into the curricula of all schools in the project, practical subjects such as industrial arts (woodwork, metalwork), home economics (domestic science), commerce and agriculture.

Stage (ii). Developing more detailed informational objectives

It is quite clear that the general objectives, while specifying what it is hoped will be achieved, give no indications of how this is to be done or even what is meant by some of the statement used. What, for example, was in the minds of the writers when they used the word 'modernize'? Or what changes were intended from the existing science curriculum in order that it should have 'greater emphasis'? It is at this stage that much more detail must be supplied, so that each of the four main areas comprising the general objectives, can be used as guides for producing the necessary specific planning objectives. As was stated in Chapter 2, page 23, the usual approach to producing specific objectives is to state them in terms of student behaviour. In the present instance, therefore, it becomes necessary to consider for example, what changes it is expected will take place in the behaviour of secondary school graduates, if the quality of their education is 'improved'. It must be equally clear that unless the word 'improve' is detailed in much more precise behavioural terms, it will be quite impossible either to know what changes should be made in the curriculum in order to bring about the improvement, or to carry out any evaluation exercise to determine whether the improvement has or has not occurred.

The exercise of determining in detail the student behavioural changes which are subsumed in the use of the word 'improve', also needs to be carried out not only for each subject of the curriculum, but also for each of the other three areas included in the General Objectives. It is quite impossible in the limited space here even to set out the range of alternative behaviours possible in each area, let alone to provide the detailed information necessary to begin the task of writing specific behavioural objectives in all the general areas listed in Stage (i). It will be noticed, however, that some overlap occurs across the General Objectives. The additional information required for producing specific planning objectives for the first two General Objectives - for 'improving' and 'modernising' the curriculum - will necessarily have much in common, and that dealing with the subject of science, will overlap with the third General Objective. Since it is an important aspect of the 'new' curricula to be introduced into the schools to be provided on this project, science will be used as a further illustrative example.

Probably the best way of providing the additional details needed to specify the requirements of the curriculum under consideration, is to write out a further set of objectives, which, in their specificity fall about halfway between the General Objectives and the final Planning Objectives. The following are examples of the kind of objective intended.(2) This list is certainly not exhaustive - the judgements which will determine in actual practice what will, or will not be included can only be determined from discussion with the World Bank experts concerned with the project, and the educational officials of the government of the country concerned.

Secondary School Science Curriculum - Informational objectives in the Cognitive and Affective Domains

1. To develop fundamental skills involved in procuring, recording, organising, interpreting and reporting scientific information, and the linguistic skills that enhance the communication of ideas.
    (a) To be able, precisely and accurately, to record observations and ideas for use in theory building and concept formation.
    (b) To use recognised library research skills to extend knowledge of the accomplishments and the unsolved problems of science.
    (c) To use appropriate mathematical tools, symbols and operations to describe, interpret and organise data to communicate ideas in a precise manner.
2. To develop basic understandings of terminology, concepts and procedures which further the process of interpreting phenomena, predicting and controlling events and generating more powerful theories.
    (a) To demonstrate knowledge and comprehension of scientific terminology and specific facts which have been emphasised in the curriculum.
    (b) To demonstrate knowledge and ways and means of dealing with specifics (knowledge of the ways of organising, studying, judging and critizing).
    (c) To be able to describe major theories and structures of science.
    (d) To demonstrate a knowledge of the relationships between science and society.

---

(2) Both these and the Planning Objectives which follow, have been adapted from a selection given in R.J. Kibler, et al. 1970, originating in A.N. Gentry, (Ed.), Planning a continuous science programme for all junior high schools Youth, Riverside, California: Office of the Superintendent of Schools, and Summary Report, Workshops on Behavioural Objectives, Washington D.C., National Science Supervisors Association and National Science Teachers Association, 1967.

3. To develop rational thinking powers which underline scientific modes of inquiry.
    - (a) To be able to generate data to support or refine theories.
    - (b) To generalise the usefulness of theories into broader concepts.
4. To develop those scientific values, appreciations and attitudes which underline the personal involvement of an individual as he investigates the environment of mankind.
    - (a) To become intrigued by phenomena in the environment.
    - (b) To exhibit enjoyment in processing data, building theories, and seeking concepts.
    - (c) To show appreciation for the rational nature of science.
    - (d) To interact with others in building theories and seeking data during laboratory investigations and reporting and discussion periods.
    - (e) To demonstrate personal integrity when reporting work.
    - (f) To respond to beauty in the environment.
    - (g) To understand that science is generated by people with a compelling desire to understand the natural world.
    - (h) To appreciate the cultural conditions under which the scientific enterprise is promoted.

By no means all the objectives listed above are written in behavioural terms. Their intent, however, is to convey in some further detail, what it is hoped the new science curriculum will achieve. It is from these that the more specific planning objectives must now be produced.

Stage (iii). Producing the Planning Objectives

This is probably the most important stage, both for the development of the desired curriculum, and for the evaluation process. It cannot be lightly undertaken, therefore, and inevitably will be time-consuming. It is suggested that the task is one best carried out by local personnel, who have undergone an appropriate training, although it may be considered desirable to have attached to the working group an expert consultant appointed by the World Bank.

Again, because of the limitations of space, a full set of Planning Objectives for the science curriculum being illustrated here, cannot be produced. One or two examples developed from the Informational Objectives listed above are given, however, in order to act as illustrations. In practice, they should cover the whole of the intended curriculum, although, as will be pointed out later, the list may be curtailed, extended or modified according to the information fed back following the various stages of the formative evaluation.

The marginal numbers refer to the Informational Objectives listed above.

1.a. When recording data obtained from observations and investigations, the students: make simple diagrams which illustrate objects, events or concepts, keep written logs of observations; record quantitative data in tabular or other suitable form; represent ideas graphically; use precise and unambiguous language to describe observations; organise recorded data in a manner that enhances their subsequent interpretation.

1.b. When obtaining information and ideas from the library, the students use card catalogue and other standard references sources; use other resources when information is not available in school or community library; document statements and opinions; consider recency of publications and competency of authors; use original sources when available.

1.c. When conducting and reporting laboratory investigations and observations, the students: use quantitative terms to describe observations; use metric measurements; use scientific notation in recording and reporting very large and very small numbers; use pictures and diagrams to indicate actual size of objects when the scale is given; graph ordered pairs of data; predict from graphs by extrapolation; infer the equation for a straight line graph; represent ideas in the form of mathematical models; apply operations of arithmetic and simple algebra to science problems and data processing.

2.a. (i) When shown a familiar diagram of a plant cell, the students identify correctly at least seven of the following parts; chloroplast, cell wall, plasma membrane, cytiplasm, Golgi apparatus, mitochondria, nucleolus, chromatin, plastid, nuclear membrane.
(ii) When asked to define a list of key terms (i.e., atom, biosphere, catalyst, diffusion, energy, substance, solution) the students state acceptable operational definitions for at least 70 per cent of the terms.

2.b. (i) When given the names of ten previously encountered organisms, the students classify at least seven of them under the proper kingdom, phylum and class.
(ii) When given previously unencountered hypotheses about environmental phenomena the students describe one or more appropriate methods for testing each hypothesis.

2.c. When asked to describe how classification of living things provides evidence to support the theory of evolution, the students state that intergrading forms of plants and animals differing from earlier species indicate that evolutionary change is probably taking place today in living things.

2.d. In a social studies class the students give examples of new technologies and/or science breakthroughs which resulted in changes in the way we live.

3.a. When provided with a microscope and yeast culture and asked to provide data that supports their theory that yeast is alive because it is made up of cells that grow and reproduce, the students observe growth and splitting-off of buds, record observations in a series of drawings, or describe the sequence of events in writing.

4.a. When seeking explanations for objects and events, the students examine objects carefully, ask relevant questions, use a combination of senses to collect data, volunteer to change the conditions in which an event took place, look for problems and not discrepancies, persist in seeking causes for events which are not immediately apparent, see and point out discrepancies which others do not see, tolerate periods of time when there is conflicting and incomplete data.

4.b. (i) During post-laboratory evaluations, the students: suggest logical explanations for events, point out contradictions between data and theory, criticize explanations which cannot be tested in a rational way.
(ii) The students seek and read books which describe the creative genius of scholarly scientific works of the past and present.

4.c. When presenting oral and written reports and writing laboratory and field data logs, the students: report data which conflict with their own theories, give credit to others whenever due, admit their mistakes, point out the limitations of their data, log only those data which they actually gathered.

4.d. When observing and investigating the environment, the students search and report evidences of order, symmetry and precision.

4.e. (i) The students demonstrate in their writing that they accept the concept that the social and economic climate will support or discourage the scientific enterprise.
(ii) The students in writing or discussion give logical arguments for or against national policies for the planning of research.

DEVELOPMENT OF APPROPRIATE PROCEDURES AIMED TO ACHIEVE THE OBJECTIVES

Although of paramount importance for the development of the curriculum, since it is not a specific stage in the evaluation process, little will be said about this actual procedure. It may be noted, however, that the very nature of many of the Planning Objectives largely determines not only the kind of instructional procedures required for their achievement, but also the circumstances in which they need to be carried out. Many objectives, for example require a laboratory setting, and hence could not be achieved if schools were not equipped with the appropriate rooms and apparatus. Others require field work, which accordingly must be built into the design of the curriculum. If the circumstances demanded for the fulfilment of some of the objectives are not available, then it may well be that even at this stage, such objectives will have to be abandoned or at least modified to allow for their achievement within the limitations imposed by the instructional procedures considered possible. Changes to the original set of Planning Objectives brought about by restrictions of this kind, form the first stage in the feedback process, which will continue throughout the formative evaluation which follows.

For the purposes of this illustration, it must be supposed that, incorporating any changes or modifications required by any possible restrictions, a curriculum is developed which is aimed to achieve the desired objectives. The production of a new curriculum is, again, no easy matter, involving as it may the writing of new textbooks, new handbooks for teachers, and the development of new materials for students. Such work is best carried out, within each specific subject, by an appropriately trained team operating in the country receiving the World Bank credit loan. It may be considered desirable to establish a Curriculum Development Unit which could receive expert help and guidance from outside personnel appointed by the World Bank or some other international organisation.

METHODS IN CARRYING OUT A FORMATIVE EVALUATION

The basic kind of evaluation required in the sort of project being described here, is, of course, formative. This will necessarily be so in any project concerned with changes in the curriculum, since its function is to ensure that objectives of the curriculum are being achieved. The need to carry out any summative evaluation will only arise if it is required, for any reason, to make comparisons of schools following a new curriculum with ones following an existing one.

The procedure to be followed in a formative evaluation involves:

(a) Devising a set of instruments aimed to measure the objectives which it is hoped are being achieved.

(b) Determining the sample to be included in the evaluation.

(c) Applying the instruments at appropriate times.

(d) Assessing and interpreting the results.

(e) Arranging appropriate feedback and planning repetitions of the evaluation procedures.

(a) Devising the instruments

Four different types of instruments may be needed in a formative evaluation, although they will not necessarily require to be administered at the same time. These are (i) Tests for students aimed to measure the achievement of cognitive objectives, (ii) Attitude scales of questionnaires designed to measure the achievement of objectives in the affective domain, (iii) Questionnaires for students, aimed to provide background and other information which will be of value in interpreting the results, (iv) Questionnaires for teachers, designed to obtain information about their background and teaching experience, and their views on the problems encountered in operating the new curriculum. It should be noted that projects which involve the introduction or development of practical subjects into the curriculum (as, for example, the General Objective 4 listed on page 70), may require the preparation of additional special assessment techniques.

The devising of all of these instruments requires a high degree of knowledge and skill and is clearly a task which can only effectively be carried out by a team which has been established and trained for the purpose. It is recognised that a problem exists in many developing countries likely to receive a World Bank loan, of finding sufficient people from within the countries who would have the qualifications and experience to become members of such a team, and it may well be a necessary adjunct of undertaking evaluation, for appropriate training courses to be established which will provide the trained personnel. Again, help from international organisations such as the IIEP or IEA, and the provision of expert consultants provided by Unesco or the World Bank, would seem essential.

For this present example, it can only be assumed that some organisation equipped with appropriately trained personnel, has been established in the country concerned. (For further comments, see Chapter 7.) With a precise set of Planning Objectives described in behavioural terms, there should be little difficulty in producing the required cognitive tests. One point only needs to be stressed here: The tests will be criterion-referenced - that is, they will be concerned with measuring the extent to which the behaviours detailed in The Planning Objectives have been achieved by all the students, and they will not be designed in order to effect a fine discrimination among the students themselves. The tests will be used to assess the effectiveness of the instructional procedures, and not of individual performances.

The evaluation of objectives in the affective domain often presents a more exacting task. There may be instances where the development of a particular attitude scale may be seen as appropriate in a formative evaluation: for example, in assessing the extent to which students have achieved the Informational Objective 2(d) listed on page 72/73, a scale measuring 'Attitudes towards science in society' might be found useful. More often, however, the information required may be better obtained from a question in a questionnaire; in the same area, for example, students might be asked questions about recent local, national or world events in which scientific discoveries influenced man's behaviour in society.

While the development of instruments for measuring students' achievements, attitudes and interests, is clearly important, there is a great deal of other information to be obtained which will have a bearing on the evaluation process. It may be both desirable and necessary to examine, for example, whether the achievement of particular objectives is dependent upon other factors present in the learning situation. How important is the age of the student, his previous educational experiences, whether he comes from an urban or rural background, the education of his parents, etc. What influence does the age, sex, training and experience of the teacher have? Is his knowledge of, interest in, and enthusiasm for, the new curriculum important? Information to provide answers to all these and other such questions is clearly necessary if the results obtained from testing the students is to be effectively evaluated.

Even extensive background information about the students and teachers is not sufficient. It is necessary to obtain further data from the teachers concerning their individual views on applying the new curriculum - what snags and problems they encountered, what parts received too much emphasis, or too little, what they thought of any new materials they had to use, or how they reacted to any new techniques of instruction. Carefully designed questionnaires to gather such information would form an essential part of the evaluation procedure.

(b) Determining the sample to be included in the evaluation

In many instances, in a formative evaluation, where the number of schools is not large, it would clearly be desirable to include all schools. The printing of tests and questionnaires, is not inexpensive, however, and where having large schools would involve the testing, scoring, processing and interpreting the results from many thousands of students, it becomes necessary to limit the number of schools in which the evaluation is carried out. Whenever this is necessary, it is important that a system of stratified random sampling is employed, in order that the results obtained can be generalised over all schools. The strata normally used would be the sex and size of the school, and a different sampling fraction would be used both for selecting schools and for selecting students within schools, in order that the over-all sampling fraction would remain a constant. Table 5-1 illustrates a typical sampling scheme.

Size of school in enrolment	(i) Fractions of schools selected	(ii) Fractions of students within selected schools	(iii)=(i)x(ii) Overall
A. Small	1/6	1/1	1/6
B. Medium	1/3	1/2	1/6
C. Large	1/2	1/3	1/6

Table 5-1. Designed sampling fractions

A sampling scheme would almost certainly be essential if a summative evaluation is to be carried out, which would involve assessments in schools other than World Bank Schools. Indeed, if comparisons are to be made it would probably be useful if a further stratification were to be made of World Bank versus 'Other' schools.

It is impossible to be precise about the point at which the number of schools becomes large enough for sampling to be necessary. Clearly it will depend both upon the financial and manpower resources available. Formative evaluation is not so much concerned with the large-scale administration of cognitive tests and complex data processing as it is with collecting a variety of other information to ascertain whether, and under what circumstances, the objectives of a new curriculum have been achieved. If the number of new schools to be built is only around five or six, with an annual intake of only 100 or so students into each, it would certainly be advantageous to include all the schools in the evaluation. The restricting factor would not be the cost of printing tests, etc. - only 1,000 or so would be required, and these could quite probably be duplicated, thus saving money - but the number of trained people available to spend sufficient time in each school to administer the evaluation instruments, and collect whatever other information may be required. Depending upon circumstances (school size, travelling distances, whether or not one or more age or grade groups would be involved) 3 to 5 schools would probably form the maximum load for any one evaluator helping in the exercise.

(c) Applying the instruments at appropriate times

The time allowed to elapse, after students have started on a new curriculum before they are tested, clearly depends upon the duration of the courses which has been designed. It may well be that those working on the curriculum development will have divided the new course into a number of units each with a specific duration time.

The actual evaluation procedures will consist of:

(i) testing the students in both the cognitive and affective domains
(ii) obtaining background information from both students and teachers
(iii) obtaining information from teachers concerning their views on the course.

The testing required in (i) can, and probably should take place at the end of each specific unit of the course, so that a full set of results is available at the end of the first year. (And each subsequent year where the evaluation is continued over a two or three-year course.) The information required in (ii) has only to be obtained once for each group of students or teachers, and could be collected at any time considered suitable during the first year, although, if it is required to assist in the evaluation of each unit of a course, it would be desirable for it to be obtained early in the year. As for the information required in (iii) this again can only be obtained after the completion of the course, or unit of the course.

In the science course that is being used as an illustration here, it might be supposed that the first year consists of three separate units, that it is intended to continue with the development of the course into a second and third year, and that each of these will also consist of a number of separate units amenable to formative evaluation. In such circumstances, the background data required in (ii) above, must be obtained during the application of the first unit, and the testing for (i), and the information required for (iii) would be gathered soon after the completion of that unit. Further testing and the collection of the views of the teachers concerned would also be obtained immediately following the end of both the second and third units. A further programme of testing and data collection would need to be undertaken at appropriate times during the second and third years of the course. At the same time, a repetition of the first year's evaluation would have to be carried out with the next group of students starting the course. It may be supposed here, however, that, following the interpretation of the data collected, and a feedback of the results to the course developers, changes will have been made, however slight, in the instruments being used to assess the achievement of the cognitive and affective objectives.

(d) Assessing and interpreting the results

Bearing in mind that the object of the formative evaluation exercise is to help improve the educational process, the main aim of the analysis and interpretation of the results obtained is to discover, whether, and under what conditions, the objectives of each unit of the course have, or have not, been achieved. The overall percentage of students achieving a given set of objectives would be determined, and these data broken down according to certain background variables obtained from the students (such as age, sex, education of parents, etc), and also from the teachers. Analyses of this kind will indicate whether, for example, there is a sex bias in an objective, or whether the results obtained appear to depend upon any of the measured characteristics of the teachers. The stark results obtained from such a data analysis will, however, need to be supplemented by the information provided by the teachers on the problems and difficulties they experienced in actually applying the course. It might be discovered for example, that, while there were no sex biases in the results relating to a particular objective, students from urban areas performed at a far less satisfactory level than those from rural farming districts. This finding might be coupled with reports from many teachers, that the achievement of the objective required some understanding of animal behaviour which many city students lacked. Again, almost complete failure in the teaching of another objective may have been revealed by the very low proportion of all students achieving any sort of success, and this could have been accompanied by a general comment from the teachers that the materials provided proved to be inadequate, or that the time allowed for the students to master the concepts involved was insufficient.

This is clearly a very important stage in the evaluation, and will necessarily require a number of judgements to be made by the evaluator. All statements should, however, be backed by the evidence obtained, and combined in a report pertaining to the particular unit covered.

(e) Arranging appropriate feedback and planning further evaluation

Ideally, reports on the evaluation of each unit of a course, should be passed back to the curriculum developers as soon as possible after the completion of each unit, and a more detailed report should be made available at the end of the first year. It is important to stress again here, that it is the success of the instruction that is being evaluated, and the course developers need to know what parts have been successful - that is, which objectives have been achieved - and which parts have not.

In general, unsuccessful instruction may have resulted from one or more of three possible causes. (i) The students concerned did not have the necessary prerequisites to enable them to achieve the desired objectives; (ii) the motivation of the students may have been inadequate for the necessary learning to take place; and (iii) the actual instructional activities, including the materials provided, were not suitably designed. Whatever the cause or causes, the curriculum designer, supplied with the evaluation reports, should now be able to effect any necessary changes. These may entail altering some of the original objectives, ensuring that the teachers concerned become aware of the need to provide some pre-instruction for some students, taking steps to help bring about an improvement in the motivation of the students, or changing the design of the course itself, perhaps by providing some additional materials.

These changes should be made so that the next year's group of students to take the new course, may follow the amended curriculum, hopefully with a much improved chance of success. The formative evaluation will, of course, need to be repeated. Some modifications may be required to the student testing instruments, but most of the questionnaires to be given to the students and teachers will remain unchanged, making this 'second round' of evaluation relatively straightforward.

While the repetition of this first year's evaluation is taking place, the procedural steps outlined above will need to be followed for the second year of the course, and in the course of time, for the third year. Whether one, or even two repetitions of the formative evaluation is required, will depend upon the complexities of the particular course, and the problems met with in its design and execution. The aim of the evaluation is, however, to ensure, that, within the circumstances pertaining in the new schools, the desired objectives are being achieved.

INDICATIONS OF THE TIME SCALE INVOLVED

It will be clear from the above account, that the processes of curriculum development and evaluation are closely inter-related, and in terms of the time scale involved, it is not a simple matter of adding on a year or so required for the evaluation to the time needed to produce the curriculum. The first stage of producing the detailed objectives, is common both to the development phase and to the evaluation. From the early specification of the General Objectives to the detailed statements of the Planning Objectives may take up to one year for a full curriculum, and this presupposes that a team of trained personnel, belonging to a Curriculum Development and Evaluation Unit are available. It is a task, however, which if started at the very conception of a new World Bank Project, should be completed by the time construction work on the new schools (or other educational establishments) is ready to commence.

The next stage consists of two aspects carried out simultaneously by two separate teams. One will be concerned with using the Planning Objectives to develop the appropriate instructional procedures required while the second will also use the Planning Objectives to construct the various tests and questionnaires required in the evaluation. Once the overall plan for the whole of a three-year course has been decided upon, the work of both teams can be concentrated upon the immediate requirements for the first year. This again may take about a year, and will be progressing during the time that the schools or other establishments are being built. This will mean that the first year of the course, and the instruments required for its evaluation, will be ready for the first intake of students. Work on the development of the second and third years of the course can be carried on during the ensuing years, as can the further development of testing and instruments and the actual process of the formative evaluation. It will be seen that, with a three-year course, and a formative evaluation extended over two cohorts or successive year groups of students, four full years will be required for the exercise to be completed. At the end of this time, however, it would be hoped that the course would be operating smoothly, and that there would be clear evidence available demonstrating that it was achieving its desired objectives.

The total Time Scale can be presented schematically as follows :

Year 1.	Conception of Project to its Final Acceptance	Development of General, Informational and Planning Objectives.
Year 2.	Building of Schools, etc.	(a) Production of overall Curriculum Development Plans (b) Development of Instructional Procedures for 1st Year (c) Construction of Evaluation Instruments for 1st Year.
Year 3.	1st Intake of Students Cohort 1.	(a) Curriculum Development for subsequent years continues. (b) Evaluation for Yr.1. Cohort 1 carried out. (c) Construction of instruments for subsequent years evaluation.
Year 4.	Cohort 1, - 2nd Year Cohort 2, - 1st Year	(a) Curriculum Development continues. 1st Year changed. (b) Repetition of Evaluation for Cohort 2, 1st Year. (c) Evaluation of Cohort 1, 2nd Year.
Year 5.	Cohort 1, - 3rd Year.	(a) Curriculum Development continued 2nd year changed.
	Cohort 2, - 2nd Year.	(b) Repetition of Evaluation for Cohort 2, 2nd Year.
Year 6.	Cohort 2, - 3rd Year	(a) Curriculum Development Completed 3rd year changed. (b) Repetition of Evaluation for Cohort 2, 3rd Year.

A NOTE ON SUMMATIVE EVALUATION

While Formative Evaluation can be seen to be an integral part of Curriculum Development, with the major aim of ensuring that the desired objectives of the Curriculum are in fact achieved, Summative Evaluation can almost be regarded as a separate piece of educational research. In the illustration that has been dealt with here, for example, a summative evaluation exercise could be carried out in Year 7. Its object would be to ascertain not only whether the new schools achieved additional objectives which were not specified in the development of their curriculum, but also to discover the extent to which the objectives of the new curriculum were achieved by other schools.

This exercise would require a different set of evaluation instruments, based both on the objectives developed for the new curriculum and those derived from the existing curriculum in the "other" schools. The first stage of the summative evaluation would therefore, consist of an analysis of the curricula of other schools to produce a set of the objectives they were purporting to achieve. From both "old" and "new" objectives appropriate measuring instruments in both the cognitive and affective domains would then be developed. These, together with additional questionnaires designed to obtain background information from students and teachers would be administered to a carefully selected probability sample of both old and new schools. The sampling design would be similar to that mentioned earlier for the formative evaluation, although the numbers of schools and students would need to be rather larger.

Because of the larger sample, and the additional information which would be gathered in a summative evaluation, the data processing would necessarily be more extensive. Providing that criterion-referenced and not norm-referenced tests were used however, the main statistic to be determined in the analyses would be the percentage of students achieving each objective, or groups of objectives. It must be remembered that the purpose of the analyses is not to differentiate students across objectives (or test items), but objectives across students. The objectives might be grouped according to the topic or section of the curriculum covered, or according to level in the classification adopted (e.g. knowledge, skills, comprehension, etc. in Bloom's Taxonomy), or according to whether they form part of the new curriculum or the old.

The results of such a summative evaluation would serve, not only as a check on the formative evaluation, but would also reveal what precise differences existed between the old and new schools in terms of what they were actually achieving. Additionally, by breaking down the analyses according to the background data collected from the students and teachers, evidence would be produced which would throw light on the factors influencing the achievement of the objectives in the varying circumstances of different school curricula, including teaching techniques and methods.

Providing a formative evaluation had first been carried out, so that much of the work producing measuring instruments had been completed, it would be possible for a summative evaluation to be completed within the space of one further year, or even less. The curriculum analyses in the "other" schools would occupy some time and manpower, but once this had been done, modifying the tests would be relatively straightforward. Careful plans would have to be made, for the administration of all the measuring instruments, their scoring, and the processing of the data collected. If it were possible to employ methods of data collection that would allow for machine scoring, the time factor could be considerably reduced.

As with the formative evaluation, much of the work, especially that concerned with the curriculum analyses, test construction and administration, could be carried out by local personnel, assuming appropriate training had been given. Unless facilities were available locally, however, the data processing would have to be carried out elsewhere, and the whole exercise would need the advice and guidance of outside experts.

A final point should be made about the carrying out of a summative evaluation on a project on which there has been no previous formative evaluation. Such situations might well arise with existing projects, or projects which had not been planned with evaluation in mind. It must be appreciated, that even if summative evaluation only is intended, it is still necessary to start with a clear formulation of the objectives it is intended should be achieved. Where a summative evaluation is required to be added to a project, as it were, it will be inevitable that the results obtained may not be everything that would be desired, especially if time does not permit the full completion of the necessary first stage of curriculum analyses. The items comprising the testing instruments may well need to be drawn from existing pools of items, which will almost certainly have been designed for norm-referenced purposes. The analyses will then have to be based on average student scores rather than objective 'scores'. While these will reveal differences which may exist between 'old' and 'new' curricula, and show the relative importance of various student and teacher 'background' variables, they will not be so helpful in pointing out weaknesses in the curricula, or in producing evidence which will lead to improvement.

CHAPTER 6

THE EVALUATION OF TEACHING EFFECTIVENESS AND TEACHER EDUCATION

The content emphasized here is appropriate in view of the importance of teachers (here used to denote part-time teachers and vocational education instructors as well as other kinds) in the educational process. In this chapter we first describe the character of objectives in World Bank Projects that refer to teaching effectiveness and teacher education. Then we deal with the five major categories of subject matter and objectives that enter into teacher education. Finally, we discuss the general character of research in these areas and its promise for future work in the evaluation of World Bank Projects.

The Objectives of World Bank Projects in the Area of Teaching

The objectives of the projects concerning teaching fall into the following main categories:

1. Increasing teacher effectiveness in general and in a specified way;
2. Increasing the knowledge and skills of teachers;
3. Improving the instruction given by teachers;
4. Improving the effectiveness of teachers as indicated by the achievement of their pupils.

Objectives in the first category can be illustrated by the following kinds of phrases, quoted from the abstracts of the objectives of projects in various countries:

> "Train instructors needed to expand and improve primary teacher training throughout the country" (Afghanistan, 1964).
>
> "Improvement and strengthening of teacher training" (Philippines, 1973).

The objectives that were classified as referring to attempts to increase the knowledge and skills of teachers can be illustrated by the following quotations from the abstracts of statements of objectives:

> "The training college expansion is to prepare more teachers for science teaching in primary schools." (Nigeria, 1964)
>
> "The training would also be designed to prepare teachers to become village leaders and would therefore include courses in community development techniques, school guidance and administration." (Cameroon, 1972)
>
> "Teacher-training programmes in a proposed teacher-training college which would be equipped to offer courses such as industrial arts, agriculture, general science and home economics." (Trinidad and Tobago, 1972)

The objectives that referred to the improvement of teaching and instruction methods used by teachers can be illustrated by the following quotations from the abstracts of their statements of objectives:

> "... staffing arrangements and teaching methods envisaged offer reasonable assurance of achieving objectives sought by the schools." (Afghanistan, 1964)
>
> "To assist and upgrade teachers and keep them abreast of curriculum development, classroom instruction will be supplemented by ETV." (Malaysia, 1971)

The basic idea that the effectiveness of teachers ought to be reflected in the achievement of their pupils was taken into account in statements of objectives that referred to such achievement as an objective to be attained through improved training of teachers. Such objectives are indicated in the following phrases quoted from the abstracts of the statements of objectives of projects in various countries:

> "The quality of primary school output improved by raising teachers' standards should provide a better secondary school input." (Gabon, 1967)
>
> "Other modifications have been designed to improve the training programmes and to extend their impact." (Chad, 1968)
>
> "A proposed teacher-training college would upgrade standards providing annually about 200 trained teachers for the improvement of the primary school programme." (Trinidad and Tobago, 1968)
>
> "This (government teacher-training college) would contribute to an improvement of the instructional programme and the quality of the output from primary schools." (Guyana, 1969)

In all their generality and brevity, these statements of objectives in the area of teaching must be considerably supplemented before they can further achievement of the goals set forth. These statements provide general and ultimate goals that need to be explicated in the form of much more specific kinds of teacher knowledge, understanding, decision-making ability, and skills in specific kinds of behaviour and performance.

Thus, in some investigations the teachers' own understanding of what they teach has been investigated. Thus, one researcher investigated teachers' understanding of arithmetic processes and concepts. It has been revealed that such ignorance on the part of teachers sometimes exists so as to throw doubt on their ability to produce understanding on the part of their pupils. Other investigations of teachers' knowledge of their own subject have revealed significant variance among physics teachers in their knowledge of the subject they were teaching, and among teachers of secondary-school mathematics. Most educators would incline to hypothesize that the more the teacher knows and understands about the subject, the better, other things being equal, his effects on his students will be.

Components of Programmes for the Education of Teachers

The conception of a teacher education programme is often regarded as including five major kinds of components:

1. The liberal education of teachers;
2. The teacher's education in the knowledge of the subject he is going to teach;
3. The foundations of education - social, psychological, philosophical and historical;
4. The theory of curriculum and instruction in the field in which one is going to teach;
5. Ability to perform the technical skills of teaching.

The liberal education of teachers is the same as that for any citizen who has received a higher education. It provides him with the knowledge and understanding of the major domains of human knowledge in the natural sciences, both physical and biological, the social sciences and the humanities. It is often regarded as valuable in itself and as something that infuses human activity in any field with greater wisdom. Just as physicists and artists, physicians and lawyers, housewives, cobblers and farmers, ought to have a certain liberal education, so should teachers.

The teacher's knowledge of his subject matter is generally assumed to be essential to his ability to teach that subject. At least, laymen tend to make this assumption. The importance of the teachers' knowledge of the subject matter may vary considerably at different grade levels and in different subject matters. It also may depend heavily on the kinds of instructional methods that are used, being more important perhaps when the lecture method is relied upon and less important when programmed instruction or independent study is involved. In any case, this aspect of teacher education programmes must be taken into account in describing and evaluating World Bank Projects in teacher education.

The foundations of education are the fields of knowledge that provide the professional teacher and educator with a grasp of the sociological, psychological, philosophical and historical concepts, principles and issues that have characterised man's concern with education over the centuries. The sociological foundations of education deal with such matters as the structure and functions of large and small social entities, ranging from whole societies to small groups.

The psychological foundations of education deal with the nature of learning and the learner - his cognitive and social development, his motivation as a state of the moment and as an enduring trait, and the nature of the achievement and its measurement and evaluation. The philosophical foundations of education deal with the relationship between curriculum and values, with the theory of knowledge underlying the curriculum, and with the proper relationship of education to the larger society and the good life. The historical foundations of education draw upon human experience down through the ages in acculturating the young so that they become useful, productive, and integrated members of the society into which they are born.

Many teachers at the beginning of their careers in the field of education have low regard for these matters, looking upon them as useless in their day-to-day interaction with their pupils. But as their experience accumulates and they begin to appreciate the problems with which they must deal, they become motivated to acquire a better understanding of the origins and nature of those problems. Then they become interested in understanding these 'foundations' of the enterprise of education. Delaying the study of these subjects until students have become somewhat experienced as teachers might make these fields receive higher regard from students in education.

The field of curriculum and instruction theory in any given subject-matter field is one that deals with the decisions about what should be taught and how it should be taught. The factors that should enter into making decisions about what should be taught are studied here. Similarly, the psychological factors - concepts and principles from the psychology of learning - are studied in the effort to approve the efficiency of teaching. Here, the teacher should learn about the ways in which characteristics of the student (his abilities and motives) should interact with the character of the educational objectives (cognitive, social-emotional, or psychomotor) and the various kinds of teaching strategies or styles that are available (the lecture method, the discussion method, role-playing, individual study, self-directed learning, humanistic methods, etc.). Studies have shown that most teachers persistently use a single method of instruction, the classroom recitation, regardless of the objectives being sought, the characteristics of their students, or the superior suitability of other methods of instruction. Teachers need to know about alternative methods, to understand their advantages and limitations, to have some basis for choosing among methods according to their desirability in various situations, and to know how to put them into effect once the decision to use such methods has been made. Then teachers will not fall into the rut that has been revealed in observational studies of classroom teaching made since the early 1900's.

Finally, we come to that part of the teacher-education programme that is intended to provide the teacher with knowledge of 'how' to perform various technical skills and decision-making processes in the interactive phases of teaching. It is in these phases that the teacher meets his students face to face. Traditionally, 'student teaching', as it is called in the United States, or 'teaching practice', as it is called in England, is used as the main avenue whereby a beginning teacher acquires these skills. During the last decade, some new approaches have appeared - 'microteaching' being especially noteworthy. Here, the beginning teacher has an opportunity to acquire these skills in scaled down situations, working with a small number of pupils (say 5) for a relatively short period (say 10 minutes), on a restricted part of the teaching process (say 'making an assignment' or 'giving an explanation').

The microteaching approach has been widely adopted in the United States and in many other countries, including Scotland, Australia, West Germany, Israel, and some African countries. It has also provided the basis for 'minicourses' which provide transportable, self-administrable, self-contained packages of teacher training products (including films, videotapes, manuals, tests, and printed instructions) designed to improve specific technical skills on the part of teachers. Among the skills for which minicourses have been developed are those related to the discussion method of teaching, techniques of developing language skills, effective questioning in a classroom discussion, tutoring in mathematics, and organizing the kindergarten for independent learning and small-group instruction. Thus far, minicourses have been demonstrated to be effective primarily in the sense of changing teacher behaviour rather than in the more important sense of improving student performance, behaviour, or achievement. But, of course, effect on teacher behaviour is a prerequisite to effectiveness in improving student achievement? And an impressive three-year follow-up has demonstrated the effectiveness of a minicourse as a vehicle for changing teacher behaviour. It was found that videotape recordings of each of 24 experimental teachers made before, immediately after, four months after, and 39 months after training revealed that the teachers were significantly above their precourse level on 8 of the 10 behaviours scored even 39 months after their training. Thus, in ability to redirect questions, prompt, clarify, avoid repeating questions, avoid repeating pupil answers, elicit longer pupil responses, reduce frequency of one-word pupil responses, ask higher-order questions (those requiring thinking), and reduce teacher talk - on all these kinds of teacher performance, the minicourse was found to be effective.

Many teacher training products have been developed in addition to the relatively expensive and elaborate minicourses, which cost approximately $1200 per course in their present form. A recent collation and analysis brought together more than 600 such teacher training products in a computerized catalogue. Each of the products was described in many different ways, such as the subject matter and grade levels, if any, for which it was specific, the kinds of teacher behaviour that it was intended to influence, the kinds of outcomes in student achievement and behaviour that the teacher behaviour was in turn intended to improve, the kinds of training situations in which the products were to be used, the kinds of facilities needed, as well as such essential information as the source, availability, and cost of the product. This computerized catalogue should be available in time for use in selecting specific teacher training products that appear to be appropriate for the improvement of teacher education in countries other than the United States, once appropriate translations and modifications of the products have been made. The products range in complexity from simple manual, simulation games, filmstrips, and programmed instructional textbooks, to more elaborate and expensive devices such as minicourses or complete modularized programmes suitable for use as entire courses at the university or college level.

It cannot yet be claimed that the teacher training product approach to the improvement of the performance aspects of teacher education has been proven successful in the sense of raising the level of student achievement of various kinds of objectives, but the approach appears to be promising enough to merit serious consideration in the development of teacher education programmes in all nations, including the less developed ones.

The foregoing enumeration of major types of components of teacher-education programmes goes well beyond the level of specificity that we have found in the references to teacher education in the World Bank Projects. It would be desirable to have specifications at at least this level if the various World Bank Projects on teaching and teacher education were to be evaluated. Otherwise, obviously, it is impossible to determine which aspect of teacher education the project is concerned with. It is also clear that, if the evaluation of a project is to be at all adequate, the level of specificity must go considerably beyond that illustrated in the foregoing description of major components of teacher-education programmes. The level of specificity illustrated above is necessary not merely to facilitate evaluation but, even more, to assist the developers of projects in formulating their own plans and objectives - that is, in deciding what they want to do.

The Research Basis for the evaluation of teaching effectiveness and teacher education

In general, the evaluation of teacher effectiveness and teacher education follows the same logic and principles as the evaluation of educational activities or enterprises of any kind. But teacher effectiveness and teacher education are involved with specific kinds of educational variables and activities and refer to relationships between variables of specific kinds. These are indicated in the diagram presented in Figure 6-1.

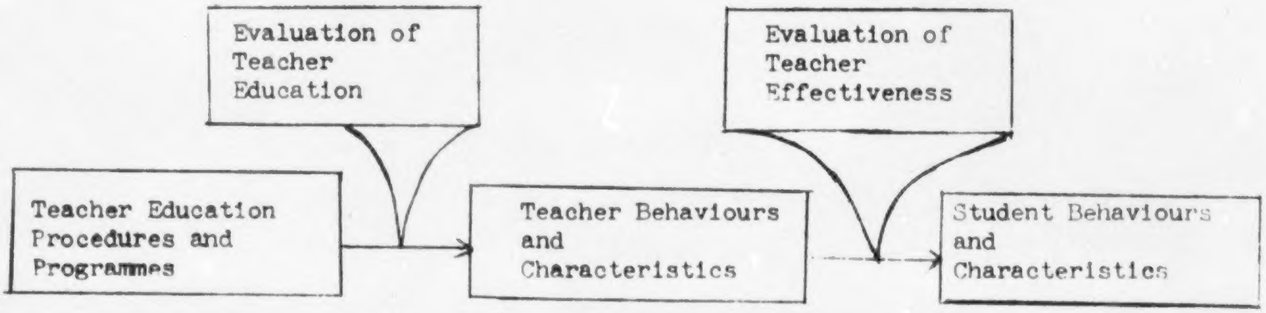

Evaluation of Teacher Education
Evaluation of Teacher Effectiveness
Teacher Education Procedures and Programmes
Teacher Behaviours and Characteristics
Student Behaviours and Characteristics

Figure 6-1. Types of Variables and their Relationships involved in the Evaluation of Teacher Effectiveness and Teacher Education

In evaluating teacher effectiveness, one studies the relationships between the teachers' behaviours and characteristics, on the one hand, and the resulting student behaviours and characteristics on the other. The teacher behaviours may take such forms as giving lectures of certain kinds, conducting recitations, giving assignments, forming discussion groups, or testing with greater or less frequency. The teacher's characteristics may consist of his age, sex, amount of experience, or knowledge of his subject.

The student's behaviour and characteristics are those with which the educational objectives are concerned. Special emphasis must be placed on determining that the relationship between teacher variables and student achievement is causal and not merely correlational.

In evaluating teacher education, one seeks to determine whether a kind of teacher education procedure - e.g. courses of various kinds, selection criteria of various kinds, microteaching, student teaching, role playing, film viewing or listening to lectures - affects the kinds of teacher behaviours or characteristics that have been set up as the objectives of the teacher-education programme.

What should be the objectives of the teacher education programme? In general, these objectives should be the kinds of teacher behaviours and characteristics that have been found to have desirable effects on students, i.e. that promote students' achievement of educational objectives. The nature of such teacher behaviours and characteristics has, of course, been the subject of endless philosophising, speculation and debate down through the centuries. In recent decades, however, a strong beginning has been made at providing answers based on scientific method, i.e. on empirical studies of relationships between teaching behaviours and student achievement. A generation of research workers has, since the 1950's, begun to amass more or less consistent findings. A recent review of such findings (Teaching behaviours and student achievement),has brought together promising results on perhaps a dozen dimensions of teaching behaviours. These results take the form of regular tendencies for a certain kind of teaching behaviour, such as 'criticism' or 'acceptance of student ideas' or 'enthusiasm' to exhibit a positive (or a negative) relationship to student achievement (adjusted for initial student ability or achievement).

In less developed countries, where World Bank Education Projects have been aimed at improving teaching effectiveness and teacher education, it is not premature to use the outcomes of this recent work. Educators can exploit the instruments - such as teacher training products,classroom interaction observation techniques, and microteaching - both for the evaluation of teaching and for its subsequent improvement. The use of these instruments almost always raises the level of teachers' awareness and understanding of their own methods, behaviour and effects.

Thus, it is intended, as the present project moves forward, to use appropriate selections of these new methods in evaluating teacher education and teaching effectiveness as these constitute areas in which the World Bank Projects have objectives.

CHAPTER 7

EVALUATION AND EXPERIMENTATION AS A COMPONENT PART OF EDUCATIONAL PLANNING

One major theme of the preceding chapters of this report has been the need to build in evaluation and experimentation procedures as a component part of all educational planning in each system of education. To recapitulate some of the examples of the types of questions where evaluation and experimentation can shed light:

- Are the revised curriculum materials or new curriculum materials 'working' with the students for whom they are intended, in the sense that the students are achieving the objectives in the particular learning materials? If not, what are the weak points in what is being attained by which students?
- Is there a systematic increase over time of student achievement in general secondary schools in science? Is this equal for all sub-groups of students in different schools in different regions within the country? If inequalities exist, where do they exist?
- Are the objectives of teacher education courses being attained by all students? If not, why not? What remedial measures can be taken?
- Are the types of teacher behaviours taught in teacher education having a positive effect on cognitive learning, affective learning, etc., in the persons taught by those teachers? If not, are there other behaviours which should be taught? If so, which?

The answers to questions such as the above are important for planners in any system of education and help towards the reallocation of resources and the way in which the various factors (input and process into the educational system) interact with each other so that improvements can be gained through the manipulation of such variables. Different models of evaluation and experimentation have already been given in the preceding chapters. We recommend that the World Bank build evaluation and experimentation into many of its future projects for the good of the recipient country (as well as the Bank) and that the Bank and the borrower be willing to allocate such resources in order that systematic design experiment and evaluation can be undertaken. In this way all parties can learn from differing educational practices. Indeed, if evaluation data exist in countries the World Bank's identification of the weak aspects in need of support will be an easier task. Furthermore, considering the large percentage of the G.N.P. and National Budget which is spent on education in most developing countries, it is almost inconceivable that little systematic effort in evaluation of education has been undertaken. Given the 'state of the art' in evaluation, it would be feasible to introduce evaluation at relatively small cost in most developing countries.

There are several examples of large-scale evaluation and experimentation in developed countries and some in developing countries. The International Association for the Evaluation of Educational Achievement has recently terminated a co-operative international project involving six subject areas at three different levels of education in each of 23 countries. Four of these countries were developing countries. The procedures for evaluation have been well tested over several decades. The quicker such evaluation and experimentation procedures can be implemented under the auspices of the national ministries of education, for the benefit of their own systems of education, the better. Just as we have shown that there are many dangers in implanting in one country a curriculum developed in another, so there are dangers in taking instrumentation for evaluation purposes from one country to another. It is the evaluators and experimenters within each country who must develop those instruments relevant to the needs of their own projects in their own cultures.

Evaluation and experimentation structures

The competencies required to undertake such evaluation work could be collected within one central unit (for example within the ministry of education) or relevant planning organisation, or within the university, or in an independent organization. The structural location within the system will depend to a large extent on the existing structures and where the competencies are currently found.

The evaluation-experimentation component would have two major functions:

(a) The systematic monitoring of the 'efficiency' of the educational system, in particular (i) student learning outcomes, measurable changes in student behaviours at various points in the educational system; and (ii), perhaps structural criteria (financial, demographic, growth and change, etc.) and societal impacts (particularly as regards student placement in the job market, changes in productivity levels, etc.).

(b) Experimentation with and evaluation of methods of teaching, curricular materials and instructional processes, methods of assessment, etc.

The monitoring of the system as a whole and the evaluation of innovations before implementation will both generate information which are of great help as a factual basis for further decision-making in the system. It is, of course, essential that the links are built and maintained between the evaluators and the policy-makers. One of the critical links in the setting up of such experimental evaluation components will be the convincing of senior officials in the appropriate ministries of the necessity of such work.

The structure, as mentioned above, will depend on the existing structures within the system.

The word research is used in Appendix V simply because it is the general term which has been used to denote, inter alia, the experimentation and evaluation aspects of educational assessment. The structures of the existing research competencies and units clearly vary from country to country. There would seem to be little use in the World Bank helping isolated individuals unless the basic number of personnel were brought together in a small unit with the 'appropriate' relationships with the Ministry of Education and Planning. The major options include: improving existing structures, creating new units with specialised functions and, in some cases, using foreign personnel for short-term periods.

In each project appraisal, the appraisal team should as a routine investigate the needs, feasibility, resources available, existing structures and propose the type of evaluation and experimentation unit to be set up or alternatively suggest a pre-investment study on the matter.

The extent to which the structure is centralized or not will be an issue. In general if no unit exists our recommendation would be to have a centralized unit placed within the ministry of education. Such a centralised unit would include structures concerned with curriculum development, teacher education, and all aspects of evaluation. We are aware of the arguments of certain researchers that the research unit should be independent of the policy-making structure but our experience does not support this view. If the policy-makers wish to take evaluation seriously they will; if not, not. The important criteria are that the unit should have effective communication with the various levels of policy-makers, the ability to collect the required data from the schools and students and the ability and wherewithal to process and analyse the data. This in turn presupposes that there will be adequate financing for a small nucleus of staff and good physical facilities. Particular projects could be added as required. The World Bank would have to support the structure such that it would continue to survive long after the World Bank had finished its work there. This will depend to some extent on the recruitment undertaken, how the career pattern fits with the structure set-up, the intrinsic interest in the work, etc.

The relationship of the evaluation and experimentation visit to the curriculum development centre, the examinations visit, and teacher training will be of prime importance to the well-functioning of the education system. All that can be done at this point is to raise the issues.

The Training of evaluators

The main aim of this section is to outline various modes of training evaluators in developing countries - an activity which is, as yet, extremely weak in these countries. The training of evaluators must be recognised, both by the Bank and the recipient country, as a priority area for technical assistance. Since, in the majority of countries, there is at least one person with knowledge of, and engagement in, evaluation, then such personnel for training can be identified, their training needs specified, and hence the modes of training planned.

(a) General situation

There is in general a lack of trained evaluators in developing countries, but the situation varies from country to country. If one takes developing countries as a whole, it would appear that the actual experience in, and knowledge of, evaluation (in all its many aspects) is limited. There are certain characteristics, when examined more closely, that do, however, point to certain variations in such experiences and knowledge among different countries. The more important characteristics are:

(i) Differences in the number of people involved in evaluation;

(ii) Differences in qualifications and the amount of experience;

(iii) Placement of educational evaluators.

In general persons with evaluation competence tend to be widely scattered in developing countries. This is associated with a lack of systematic and co-ordinated evaluation work.

(iv) Differences in experimentation and evaluation activities.

All these different characteristics can be identified for each country and will help determine the type of training programmes required.

(b) Possible modes of training

The needs of persons to be trained will differ somewhat from country to country and certainly over time. One should perhaps think of training at two levels: simple and complex. There are certain routine aspects of evaluation (e.g. the one given on curriculum evaluation in Chapter 4 of this report) which are elementary. Other forms require a great deal of knowledge in research design, sampling statistical analyses, etc. which must be known by the evaluation leaders in each country. Unfortunately, those with more advanced knowledge (but with little experience) will often 'use a sledgehammer to crack a nut'. This syndrome must be avoided.

Let us consider the following training models:

1. On-the-job training;
2. Tailor-made international intensive training courses;
3. Regional and national training courses;
4. Short-term academic training;
5. Study leaves.

1. On-the-job training

One of the quickest ways of improving the evaluation competency of a group is that they have to accomplish certain tasks within a certain time limit, but where the tasks are very specified (e.g. construct a test to measure the science achievement of 8th graders, or draw a probability sample of pupils aged 10 in the public system of education). The IEA co-operative form of research has increased competency considerably in those research centres which were relatively weak at the beginning of the IEA project.

Each national researcher had to go through a series of steps: identification of target populations, content analyses of textbooks, item writing, pre-testing of items on specially selected judgement samples, item analysis, drawing of probability samples of schools and students in target populations, identification and measuring of critical variables in society, translation and printing of all tests, briefing of testers, data collection, coding, data recording, interpretation of analyses, etc. In some cases, IEA national personnel had extremely little experience of each of these steps, and their level of competence increased dramatically.

It is difficult to compare the costs of this type of training with another. The whole project took some seven years with the bulk of the national work being done in the first 4 to 5 years. Each national centre had to pay its own staff and the data collection within its own country. IEA International paid for the international meetings of the national technical officers from all participating countries.

The important aspect is that each national technical officer (and,in turn, his assistants) had to carry out in detail each step in the evaluation process for the national system in his own country.

Another form of on-the-job training is to have a specific project begun, and for the World Bank to supply the appropriate technical assistance either for the period of the project or at regular intervals throughout the project. The technical assistants would have to be hand-picked. The selection mechanism would, we think, pose a problem for the World Bank. What is important in this type of training is that the group of evaluators remains the same from the conceptualisation of the project through to the interpretation of the results and write-up. One difficulty of this latter type of training is that for more complex data the computer facilities and/or programming personnel are not available in developing countries whereas in the former type the data processing would be undertaken internationally. Furthermore, for large-scale data collection, mark-sense cards are the most reliable tool to use and again developing countries do not have the equipment for either producing or reading them. Admittedly, labour intensive coding, punching and verifying methods can be used but these are more open to error.

In each case the bulk of the financing should come from the individual nation.

2. Tailor-made international intensive courses

Intensive courses at an international centre can be very useful provided that the course is geared closely to the needs of the participants. Several examples exist of persons with virtually no knowledge or experience of evaluation attending international courses and upon return to their countries (also without educational evaluation) initiating evaluation structures which within two or three years became the mechanism for undertaking all work in system monitoring and the evaluation work of the curriculum centre.

There is no reason why this type of result cannot occur in developing countries. In many ways it should be easier with somewhat less tradition to be overcome.

Somewhat similar courses have bean run in the past under the general headings of Seminars on Education and the Learning Process (SELEPS). Additional information on these courses can be obtained from IIEP. For some purposes universities and testing agencies run important training courses in different aspects of evaluation.

The total cost of such courses can run from $10,000 (e.g. IIEP course, where the premises cost nothing and the air fares and per diems are paid by the individual countries or by international agencies, e.g. Ford Foundation or UNDP) to $90,000 (35 persons) where all is paid by the sponsoring agency.

3. Regional and national training courses

Again, these courses should be tailor-made to meet the needs of the participants. The advantage of these regional courses is that they will be cheaper and the evaluators in a region will get to know each other's work and can set up a network collaboration. Such courses could be of 4-6 weeks' duration with as many faculty as possible being drawn from the region itself. IIEP has recently sent a document to curriculum centres in English-speaking Africa, after each centre had been approached individually concerning its own training needs. This is an example of one link in the chain of setting up such regional courses. There is every reason to suppose that this type of course is needed in different aspects of evaluation. Information on the need identification process can be obtained from IIEP.

4. Short-and long-term academic training

a) Short-term

In several countries it will be beneficial to identify certain persons who will become leaders in evaluation in their own countries and offer them training in depth. Such personnel should be selected in terms of their interest (e.g. teacher effectiveness evaluation, curriculum evaluation, survey evaluation, advanced statistical analysis, etc.) for the formation of groups of some 10-25 persons from various countries. Each group would have one or two intensive courses of three or six months' duration at one or two selected academic institutions.

It is rare for any one academic institution to be strong in all aspects of measurement, evaluation, and statistical analysis. Hence, in some cases it would be wise to have a group attend one university for one course, to be followed at another university by another course. The trainees thus receive the best training possible. It should also be possible to attach such trainees for 2-6 months to ongoing evaluation projects as a 'practical study course'. The use of carefully selected R & D centres could encompass both theoretical and practical training for such groups, provided the R & D centres were prepared to offer special courses.

b) Long-term

It has been common practice in the past decades for evaluation specialists to attend universities in other countries for a three or four-year period for their graduate work, there being no possibility of graduate work in their own country. In many ways it would be of more practical value to have students train at a foreign university for short periods interspersed by longer periods of field work in their own countries.

To this end it is anticipated that one or two universities will be selected which can offer programmes of three to four months' duration each year for a period of three to four years. Although these programmes could lead to advanced degrees, the main thrust of the work would be found in the country of origin.

5. Study leave abroad

For certain senior evaluators it would not be necessary to have further academic training, but it would be desirable for them to gain knowledge on how the problems they are facing are dealt with in practice in other countries. It is suggested that individuals would have study leave of a few weeks' duration to visit several evaluation centres known to have made progress in the particular areas of immediate importance to the centre concerned in order to observe the methods of evaluation elsewhere.

6. Levels of expertise and existing resources

A distinction was made earlier between simple and complex types of evaluation. Perhaps the simpler type should be learned by on-the-job training and national regional seminars, the more complex being reserved for other forms of training. The more complex training would have to be more subject-matter anchored, e.g. teaching effectiveness research would have to be a specialised course. It is difficult to see the borderline between the two clearly at this point, but this should become clearer as experience with training proceeds.

It should be reiterated that the World Bank could benefit from the ongoing work of ICO (International Curriculum Organisation) and IEA (International Association for the Evaluation of Educational Achievement). ICO will, in the near future, begin regional seminars in curriculum development and evaluation in which curriculum evaluation will be a special course. The first regional seminar (for English-speaking Africa) will take place in July/August 1975. The Asian seminar will follow in 1976. IEA is currently contemplating some repeat national surveys of educational achievement.

It would be useful for the World Bank to encourage some selected developing countries to join in this work, hence ensuring training on the job.

In any case the above methods of training should be examined and certain selected for experimentation. Unesco, and the World Bank, should, during PIM and appraisal missions, identify the needs and persons for the different types of training mentioned above.

Evaluation costs of typical projects

Throughout this report no reference has been made to the costs of carrying out any one evaluation of a particular project. Clearly the costs will vary from project to project, dependent upon local conditions, and the size and complexity of the project. Below we have described two different types of evaluation of two different projects both of which might be regarded as typical. We show the timetable, the competencies required and the personnel man-months.

PROJECT 1 - SUMMATIVE EVALUATION

Let us assume that a summative evalution is required of a specific vocational education programme which is operating in 500 schools in a developing country. The curriculum would be well established but the test constructors would have little experience of the vocational subject curriculum and no experience of developing tests in that subject. Testing would be carried out in some 40-50 schools with about 2,000 students in the sample involving data collection on input process and output data. This would yield a considerable volume of data. The analyses will include multivariate analyses and hence it will not be convenient or appropriate to have the data analysed by hand. It is recommended that a computer be used, and this implies periodic access to a relatively small computer. (As a general point, it should be stated that small computer facilities already exist in developing countries and often they are under utilized. The critical consideration is that the hardware is powerful enough to enable programming in a high level language such as FORTRAN or ALGOL. This enables evaluation units to develop their own computer programmes, and to take advantage of the many packages created to do this sort of work in various centres around the world).

One member of the evaluation unit will act as project director.

The timetable will be as follows:

	Time in months
Construction of tests and instruments (including pre-testing)	Month 0
Draw sample	
Printing of tests	
Administration of tests	Month 6
Coding and recording of data	Month 7
Statistical analysis	Months 8-9
Interpretation of results and write up	Months 10-12

The competencies required will be:

a) test and scale construction

b) a knowledge of sampling theory

c) statistical analysis

d) data processing

The persons required together with the time in man-months will be:

Professional

a) 1 test constructor (2 mm)

b) 1 curriculum consultant (2 mm)

c) 1 scale and questionnaire constuctor (2 mm)

d) 8 test administrators (2 mm)

e) 1 data processor (8 mm)

f) 1 statistician (1 mm)

Other

g) 2 coders (1 mm)

h) 1 secretary (12 mm)

PROJECT 2 - FORMATIVE EVALUATION OF CURRICULUM MATERIAL

The plan of formative evaluation is based on the assumption that a curriculum team prepared a draft of new instructional material. This draft has been produced in sufficient number of copies and put to try-out in a sample of approximately eight classes selected on judgemental basis in a way that it represents the student population for which the programme has been prepared.

The responsibility of evaluators will be to collect data during the try-out of the material, to summarise the data and to formulate suggestions for the modification of the programme, if needed.

The timetable of such activity is bound to the timetable prescribed for the try-out of the material. It is assumed that a course will be taught during a full school year, i.e. during ten months. Evaluation activities should start two months before the actual try-out starts, and continues during ten months of the teaching period, and then it will continue for two additional months after finishing the try-out.

The plan is prepared for a course which is taught during a whole school year to an extent of five periods in a week. Such an extent of a programme corresponds to a one-year programme in major subjects taught in school, such as mother tongue, science or practical agriculture.

../..

(Project 2 continued)

STAFFING

The task	Persons needed	Competency
1. Constructing 7 achievement tests, each one covering material of 25 periods (approximately 7 x 40 items)	3 mm	Teacher, specialist in subject-matter; some in-service training by the co-ordinator teacher.
2. Visiting schools, observing classes, 5 visits in each class, together 40 visits	2 mm	Teacher
3. Administering examinations, coding, summarising results, examining work-sheets, etc.	5 mm	Student of psychology or trained clerk
4. Co-ordinator; supervision of work, counselling, contact with schools and with writing team	4 mm	M.A. in educational psychology
5. Typing, secretarial assistance	14 mm 2 mm 16 mm	Clerk

BIBLIOGRAPHY CONTAINING FURTHER DETAILS OF SEVERAL OF THE IDEAS PRESENTED IN THE REPORT

Books and Documents

BLOOM, B.S. *et al.*, *Handbook on formative and summative evaluation of student learning*, New York, McGraw-Hill, 1971, 923p.

BLOOM, B.S. *et al.*, 'Taxonomy of educational objectives', *Vol. Handbook I: Cognitive domain*; *Handbook II: Affective domain*, New York, David McKay Co., 2 volumes, 1965.

BRUNER, J.S., *Toward a theory of instruction*, Cambridge, Massachusetts, Harvard University Press, 1966.

CAMPBELL, Donald T., STANLEY, Julian C., *Experimental and quasi-experimental design for research*, Chicago, Rand McNally, 1966.

CRONBACH, Lee J., *Essentials of psychological testing*, New York, Harper & Bros., 1970, 3rd edition.

EBEL, R., *Encyclopedia of educational research*, Ontario, Collier-MacMillan, 1969, 1522p. (4th edition).

EBEL, Robert L., *Essentials of educational measurement*, Englewood Cliffs, NJ, Prentice-Hall, 1972, 622p.

GAGE, N.L. (ed.). *Handbook of research on teaching*; a project of the American Educational Research Association, Chicago, Ill., Rand McNally, 1963.

GAGNE, R.M., *The conditions of learning*, London, Holt, Rinehart & Winston, 1970, 2nd edition.

GROBMAN, H., *Evaluation activities of curriculum projects; a starting point*, Chicago, Rand McNally, 1968. (American Educational Research Association, Monograph series on Curriculum evaluation No. 2)

GUILFORD, J.P., *Psychometric methods*, New York, McGraw Hill, 1954.

HAVIGHURST, Robert J., *Developmental tasks and education*, New York, David McKay, Inc., 2nd edition, c.1948, 100p.

KISH, Leslie, *Survey sampling*, New York, Wiley, 1965.

LINDQUIST, E.F., (ed.), *Educational measurement*, Washington DC, American Council on Education, 1951.

PIDGEON, Douglas, YATES, Alfred, *An introduction to educational measurement*, London, Routledge and Kegan Paul, 1968.

ROSENSHINE, B., Teaching behaviours and student achievement, Slough, Bucks, National Foundation for Educational Research in England and Wales, 1971.

SUCHMAN, E.A., Evaluative research, New York, Russell Sage Foundation, 1967.

THORNDIKE, Robert L. (ed.), Educational measurement, Washington DC, American Council on Education, 1971, 2nd edition.

THORNDIKE, Robert L., HAGEN, Elizabeth, Measurement and evaluation in psychology and education, New York, Wiley, 1969, 3rd edition.

TYLER, R.W., Basic principles of curriculum and instruction, Chicago, Ill., University of Chicago Press, 1949.

TYLER, R.W., GAGNE, R.M., SCRIVEN, M., Perspectives of curriculum evaluation, Chicago, Ill., Rand McNally, 1967. (American Educational Research Association Monograph series on curriculum evaluation No. 1)

YATES, Alfred (ed.), Grouping in education, A report sponsored by the Unesco Institute for Education, Hamburg; Stockholm, Almqvist and Wiksell, New York, Wiley, 1966.

Periodical articles

BLOOM, Benjamin S., 'Innocence in education', School Review, Vol. 80, No. 3, May 1972, p.333-352.

BLOOM, Benjamin S., 'The role of the educational sciences in curriculum development', International Journal of Educational Sciences, No. 1, 1966, p.5-66.

CARROLL, J.B., 'A model of school learning', Teachers College Record, Vol 64, No. 8, May 1963, p.723-733.

CRONBACH, Lee J., 'Course improvement through evaluation', Teachers College Record, Vol. 64, No. 8, May 1963, p.672-683

KRATHWOHL, D.R., 'Stating objectives appropriately for program, for curriculum, and for instructional materials development', Journal of Teacher Education, Vol. 16, No. 1, March 1965, p.83-92.